Université de Montréal

The role of evolution in the genetic susceptibility of intracranial aneurysm

Par Sirui Zhou

Programme de Sciences Biomédicales
Faculté de Médecine

Thèse présentée à la Faculté des Études Supérieures
En vue de l'obtention du grade de Ph.D.
En Sciences Biomédicales

Décembre 2016

RÉSUMÉ

Les Inuits du Nunavik regroupent des peuples autochtones de l'Arctique qui au fil de l'histoire ont formé une petite population isolée dans la région du Nunavik (nord de la province de Québec, Canada). Le profil génétique unique des Inuits du Nunavik est le résultat d'une adaptation à leur milieu de vie et il est considéré comme lié à certaines de leurs susceptibilités pathologiques. Une évolution neutre, ainsi qu'une suite d'événements adaptatifs, ont façonné le génome de ces Inuits et indirectement engendré leur prédisposition accrue à certains troubles cardio-vasculaires et cérébro-vasculaires (ex : hypertension et anévrismes intracrâniens (AI)).

Les AI sont des faiblesses cérébro-vasculaires localisées pouvant mener à des dilatations et renflements localisés de la paroi vasculaire. De telles distorsions sont susceptibles de perturber les vaisseaux et entrainer des hémorragies sous-arachnoïdiennes. Les AI sont un désordre complexe dont la prévalence est élevée (4-8%) et différentes populations (non reliées aux Inuits) ont aussi un risque accru de développer des AI. Le développement des AI est associé à la fois à des facteurs environnementaux et génétiques; plusieurs études génomiques ont identifié des régions associées aux AI. Une grande part de l'héritabilité génétique des AI demeure encore inexpliquée, en particulier dans des populations autres que les finnois et les japonais. Toutefois il est à noter que peu d'études génétiques des AI ont tenu compte de la contribution de variations génétiques spécifiques à la population étudiée.

Pour améliorer nos connaissances sur la part encore inexpliquée de l'héritabilité des AI (qui implique une grande hétérogénéité génétique et des variations peu pénétrantes), nous avons combiné le séquençage à haut débit au génotypage des polymorphismes sur puces afin d'établir la signature génétique des deux populations fondatrices du Québec prédisposées aux AI (Inuits du Nunavik et Canadiens français). Comme ces populations ont des caractéristiques distinctes, nous avons utilisé des approches différentes pour tenter d'identifier des facteurs de risque génétiques. Les Inuits du Nunavik représentent une population autochtone et de nombreux aspects de leur signature génétique diffèrent de celles des principales populations, nous avons choisi d'approfondir leur histoire et profil génomique avant de vérifier si des associations génomiques pouvaient être établies avec les IA. Nous avons tout d'abord examiné les régions codantes du génome et observé de nombreuses composantes génétiques spécifiques aux Inuits

du Nunavik qui reflètent que la population s'est adaptée à son environnement (ex: ascendance Inuit homogène, augmentation du déséquilibre de liaison et signature génétique). Des signes de sélection naturelle, jusqu'alors non-identifiés, ont révélé une accumulation de variations génétiques dans des gènes impliqués dans le processus d'adhésion cellulaire et de la réponse immunitaire (ex. *CPNE7* et *ICAM5*). D'autres analyses ont révélé un variant dans le gène *CCM2* qui présente une sélection positive et est significativement associé aux AI chez les Inuits du Nunavik. En ce qui concerne l'étiologie génétique des AI dans la population des Canadiens français, nous avons adopté une approche différente et utilisé des variations spécifiques aux Canadiens français, qui ont été identifiées par le séquençage complet de l'exome. Ces variations ont permis de générer une liste de gènes à risque potentiel, qui ont ensuite été priorisés en utilisant un test d'association par gène de type « burden ». *RNF213* est apparu comme le meilleur gène candidat; il est vraisemblablement la conséquence d'une dérive génétique. Les mesures génétiques et fonctionnelles subséquentes ont validé la contribution possible de *RNF213* au développement des AI chez les Canadiens français. Les résultats présentés dans cette thèse soulignent l'importance de prendre en compte le contexte génétique spécifique apporté par l'évolution lorsqu'une maladie complexe est étudiée. Il a également été montré que les variations d'un gène spécifique (ex. *CCM2* et *RNF213*) peuvent contribuer au développement de différentes pathologies lorsqu'ils sont observés dans des populations distinctes. D'une manière générale, nos découvertes génétiques ont permis d'identifier de nouvelles « pièces » génétiques et pour avancer le « casse-tête » incomplet de l'héritabilité génétique des AI ; la génétique des populations a été un élément clé pour cette avancée.

Mots-clés : Inuits du Nunavik, Canadiens-français, anévrismes intracrâniens, séquençage complet de l'exome, adaptation, maladies cérébrovasculaires, génétique des populations

ABSTRACT

Nunavik Inuit is a group of Arctic indigenous people, who have historically regrouped as a small and isolated population across the Nunavik region of northern Quebec (Canada). The unique genetic profile of Nunavik Inuit is the result of years of adaptation to their living condition, and it is likely responsible for their increased susceptibility to certain pathological conditions. Prior studies have shown that as a consequence of neutral evolution or past adaptive events, today's Inuit are predisposed to cardio-cerebrovascular disorders, e.g. hypertension and intracranial aneurysm (IA).

IA is defined as localized cerebrovascular weakness which leads to vascular dilation or ballooning, and such distortions are susceptible to disrupt the affected vessels and lead to subarachnoid hemorrhage. It is a complex disorder with a high prevalence (4-8%) and certain populations have been observed to present an increased risk of developing IA. Both environmental and genetic factors are deemed to contribute to the development of IA and in regards to the latter, independent genome-wide association studies (GWAS) have identified multiple loci associated with IA. Nonetheless, there is still a large portion of the genetic heritability of IA, especially in different populations other than Finnish and Japanese that remains unexplained. However, fewer IA genetic studies have taken in consideration the contribution of population specific genetic variants.

To address some of the IA missing heritability that is deemed to be accountable to its genetic heterogeneity and low penetrance, we have combined high throughput sequencing (HTS) with SNP-chip genotyping to examine the genetic signatures of two founder populations from Quebec that are predisposed to IA, including Nunavik Inuit and French-Canadians (FC). Because these populations have distinct genetic characteristics, we used different approaches for the identification of genetic risk factors. Nunavik Inuit is an indigenous population and many aspects of its genetic signatures differ from those of separate world-wide major populations; therefore we chose to conduct extensive population genetic studies in regards to their genetic history and genomic profile before we undertook to test if any association could be established between genomic loci and disease susceptibility. We observed many genetic components that are specific to the Nunavik Inuit population, including its homogeneous Inuit ancestry, increased

linkage disequilibrium (LD) and genetic signatures which reflect the population had a long history of adaptations to their environment. Previously unidentified signals of natural selection, which focused on coding regions of the genome revealed an accumulation of genetic variants in genes involved in the processes of cell adhesion and immune responses (e.g. *CPNE7* and *ICAM5*). Further analyses revealed a variant in *CCM2* to be under positive selection and significantly associated with IA in Nunavik Inuit. In regard to the genetic etiology of IA in the French-Canadian population, we took a different approach and used French-Canadian specific variants that were identified by whole exome sequencing to generate a list of potential risk genes; which were further prioritized using a gene based burden association test. *RNF213* emerged as a prime candidate gene that had undergone possible genetic drift and the follow-up genetic and functional examinations further supported its potential contribution to the development of IA among French-Canadians. The results presented in this thesis highlighted the importance of taking into consideration the specific genetic background brought by natural selection or genetic drift, both are driving forces of evolution, when a complex disease is being studied. It also further confirmed that variants in a specific gene (e.g. *RNF213* or *CCM2*) may contribute to the development of different pathogenesis when examined in distinct populations. Overall as our genetic findings identified new genetic "pieces" that further completed the missing heritability "puzzle" of IA genetics; evidence for these "pieces" were interestingly highlighted through population genetics.

Keywords: Nunavik Inuit, French-Canadians, intracranial aneurysms, whole exome sequencing, adaptation, cerebrovascular disease, population genetics

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1KGP	1000 genome project
AAA+	ATPases associated with diverse cellular activities
ABC	approximate bayesian computation
ATPase	adenosine triphosphate hydrolase
AVM	arteriovenous malformation
BCE	before Common Era
BP	before present
BWA	burrows-wheeler aligner
CADD	combined annotation dependent depletion
CAND2	cullin associated and neddylation dissociated 2
CCDC80	coiled-coil domain containing 80
CCM	cerebral cavernous malformation
CD/CV	common disease/common variants
CDKN2BAS	cyclin dependent kinase inhibitor 2B antisense RNA 1
CE	Common Era
CEU	Utah Residents (CEPH) with Northern and Western Ancestry
CHB	Han Chinese in Bejing, China
CNV	copy number variations
CPNE7	copine vii
CPT1A	carnitine palmitoyltransferase 1A
CTA	computerized tomography angiography
CVD	cardiovascular disease
EC	endothelia cells
ECM	extracellular matrix
EDNRA	endothelin receptor type A
FADS	fatty acid desaturase
FBAT	family based association test
FC	French Canadians
FHIT	fragile histidine triad

FIA	familial intracranial aneurysms
F_{ST}	fixation index
GATK	genome analysis tookit
GCTA	genome-wide complex trait analysis
GEO	gene expression omnibus
GI	Greenland Inuit
GO	gene ontology
GRCh37	genome reference consortium human version 37
GREML	genomic-relatedness-matrix restricted maximum-likelihood
GRS	genetic risk score
GST	glutathione S-transferase
GTEx	genotype-tissue expression
GWAS	genome-wide association study
HC	high-confidence
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HTS	high-throughput sequencing
HUVECs	human umbilical vein endothelial cells
HWE	Hardy–Weinberg equilibrium
IA	intracranial aneurysm
IBD	identical by descent
ICAM5	intercellular adhesion molecule 5
IGV	integrative genomics viewer
JPT	Japanese in Tokyo
LCL	lymphoblastoid cell lines
LD	linkage disequilibrium
MAF	minor allele frequency
MCA	middle cerebral artery
MCMC	Markov Chain Monte Carlo
MDS	multidimensional scaling
MMD	moyamoya disease
MMPs	matrix metalloproteinases

MRA	magnetic resonance angiography
mtDNA	mitochondrial DNA
NAT	Native Americans
NHLBI GO-ESP Sequencing Project	National Heart, Lung, and Blood Institute Grand Opportunity Exome
NUI	Nunavik Inuit
NUIrelate	populations have close common ancestors with Nunavik Inuit
NWT	Northwest Territories
OD	optical density
PBS	population branch statistics
PBS	phosphate-buffered saline
PCA	principal component analysis
PMSF	phenylmethane sulfonyl fluoride
POLR2A	RNA polymerase II polypeptide A
PUFAs	polyunsaturated fatty acids
QC	quality control
qRT-PCR	quantitative real-time polymerase chain reaction
RAF1	Raf-1 proto-oncogene, serine/threonine kinase
RNF213	ring finger protein 213
ROH	runs of homozygosity
RQ	relative quantification
SAH	subarachnoid hemorrhage
SD	standard deviation
SE	standard error
SFS	site frequency spectrum
SIA	sporadic intracranial aneurysms
SKAT	sequence kernel association test
SNP	single-nucleotide polymorphism
SNV	single-nucleotide variation
SOX17	sex determining region Y box 17
STARD13	star-related lipid transfer domain containing 13

STAT2	signal transducer and activator of transcription 2
T2D	type 2 diabetes
TBC1D4	TBC1 domain family member 4
UTR	untranslated region
VCF	variant call format
VSMCs	vascular smooth muscle cells
VT	variable thresholds
WES	whole exome sequencing
WGS	whole genome sequencing
WSS	wall shear stress
WSSG	wall shear stress gradient
YRI	Yoruba in Ibadan, Nigeria

Something ends, something begins.

--Andrzej Sapkowski

ACKNOWLEDGMENTS

First of all, I would like to give my thanks to both of my supervisors, Drs. Guy Rouleau and Lan Xiong, whose mentorship have guided me all these years and allowed me to evolve and become what I am today. Thank you for giving me the opportunities and providing me with the platform for taking this initial step on the path to become a scientific researcher, and also for encouraging me to pursue any project in an unconventional direction. What I gained from you is not only knowledge itself, but the art of thinking, self-learning, and making decisions, which I shall always remember and practice in my future career and life.

I have benefited greatly from this healthy research environment, for which I would like to credit the wonderful team that I am lucky to work with: Dr. Patrick Dion, who has always been there and helped me even with the smallest problems; and to Amirtha for your special companionship and dedicated friendship for the past six years; to Pingxing and Qin for our talks and discussions; to Alex and Dan who helped me a lot in the bioinformatics which brought me closer to this field; to Daniel and Amélie who provided professional, essential helps in my experiments; and to Helene, Pascale, Simon, Cynthia, Claire, Gaby, Ziv, Hussein, Loubna, Sandra, Alanna, Faezeh and to all the present and past lab members from the Rouleau lab, I would like to express my sincere appreciations for your help and accompany throughout my Ph.D. life.

I wish to also thank Dr. Simon Gravel for lending me a hand with his expertise in population genetics and Dr. Akio Koizumi who helped us to validate one of the major findings presented in this thesis. And also to my committee members Drs. Marie-Pierre Dubé, Nicolas

Dupré and Daniel Lajeunesse, thank you all for your support and valuable suggestions toward my research.

The works presented in this thesis would not be possible without the support from the joint scholarship of China-scholarship Council and Université de Montréal and the scholarships from Université de Montréal la Faculté des études supérieures et postdoctorales, and Canadian Institutes of Health Research (CHIR).

Finally, I want to thank all my friends in Montreal and back in China, who have made my life fuller; and my family, to my parents, Ping and Li, without whom I would not be the person I am today, who would always support my decisions and tell me not to worry. And special thanks to David Conroy and his family, for perpetual support and providing inspiration in so many ways. It is nice to have you all for perhaps the most memorable and productive years in my life, and I look forward to the years coming that we would all to accomplish much more.

CHAPTER 1: INTRODUCTION

1.1: PREFACE

Gaining an understanding of genetic profiles across the full length of genomes has proved to be critical to our understanding of human genetic history and local adaptation; it has concomitantly helped to delineate the genetic etiology for a number of complex human disorders. The recent advancements made for high-throughput sequencing and genotyping technologies have greatly enhanced our present ability to identify the genome-wide genetic signatures of specific populations, as it eased the unravelling of disease risk factors. The literature reviews presented in the introduction section will cover two seemly unrelated area of studies, nonetheless both areas are essential for the question to be explored in this thesis: 1) recent advances of the population genetic studies about Inuit and related Arctic indigenous populations; and 2) genetic studies of IA, a complex cerebrovascular disorder with increased prevalence in certain populations including French-Canadian and Inuit.

Chapter 1.2 will essentially be dedicated to findings recently published by a number of genetic and genetic anthropological studies of Greenland Inuit and indigenous Siberians. This chapter will provide details about the genetic history, local adaptations and predisposition to diseases of these Arctic indigenous inhabitants. The chapter will also highlight the importance of disease gene discovery through historically small and isolated founder populations.

Chapter 1.3 will review and discuss a number of hypotheses regarding IA etiology by highlighting recent findings from studies that aimed to identify IA predisposing genetic risk factors. In regard to this last element the chapter will emphasize the importance of whole exome

sequencing (WES) and genome-wide association studies (GWAS). Interestingly, a number of recent large-scale IA association studies were conducted using IA predisposed populations (e.g. Finnish and Japanese), these raised points that are critical when addressing the relationships between complex diseases and population-wide genetic signatures, which have been introduced throughout the process of human evolution.

1.2: THE GENOMIC LANDSCAPE AND DISEASE RISKS OF INUIT PEOPLE

Manuscript: The genetic history, adaptation and disease predisposition of Inuit and Arctic peoples

Authors and Affiliations:

Sirui Zhou,[1,2] Patrick A. Dion,[1,3] Lan Xiong [2,4], Guy A. Rouleau[1,3,*]

[1]Montreal Neurological Institute and Hospital, McGill University, Montréal (Que), Canada.

[2]Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal (Que), Canada.

[3]Department of Neurology and Neurosurgery, McGill University, Montréal (Que) Canada.

[4]Centre de recherche, Institut universitaire en santé mentale de Montréal, Université de Montréal, Montréal (Que) Canada

Abstract

The arctic aboriginal peoples, especially Inuit, are interesting for their way of life and for being the last branch in the "Out of Africa" migration. They are also noteworthy for their distinctive genomes, their adaptation to a distinctively harsh environment, and their characteristic disease predispositions. This review will provide the emerging conclusion regarding a series of recent genetic studies that examined Inuit or other present or ancient arctic populations. It will summarize what has been established about the Inuit population, and also highlight critical elements for future research directions based on reconstruction of population history, discoveries of genetic signatures

in positive selections, and population specific disease profiles associated with specific founder variations.

Introduction

According to Reich *et al.*[1], three-waves of migration populated the Americas starting 15,000 years ago through the Bering Land Bridge: First Americans, which most of the Native Americans descended from; second, the Eskimo-Aleuts from the arctic, with half of their ancestries from the second stream of Asian gene-flow; and Saqqaq-NaDene speakers, which partially consists of the third stream of gene-flow.

The arctic circumpolar regions are inhabited by many indigenous people (**Fig 1**), such as the Siberians, Samis, Indians and Inuit. The Inuit originated from East Asia and crossed the Bering Strait as part of a final wave of migration about 6,000 years ago, with the present population have an approximate 170,000 individuals in the circumpolar region. Arctic Inuit populations, most of the Inuit population, live in Canada, Greenland and Alaska[2], namely Greenland and Canadian Inuit, Alaska's Yupik and Iñupiat, and Siberian Yupik[3].

Throughout the years of their inhabitation, especially after 1950s, Inuit populations have undergone changes that led them to abandon their traditional life style. The diet of modern Inuit now contains more carbohydrate than their traditional diet which was based on unsaturated fat and protein to fulfill their energy requirements. In recent history the climate of the Canadian Arctic has been undergoing through drastic changes and these negatively impacted the health of some of its inhabitants[4]. Indeed a higher incidence of complex diseases has become noticeable (e.g. cardiovascular disease[5], infectious diseases[6], and even cancer[7]). Metabolic and cardiovascular

diseases are now a major public health concern for in the 21st century indigenous Arctic populations[5]. This review aims to systematically discuss the genetic history of Inuit and related Arctic peoples, as well as to examine how their distinctive genetic architectures contribute to their special metabolic and physiological traits and disease predispositions.

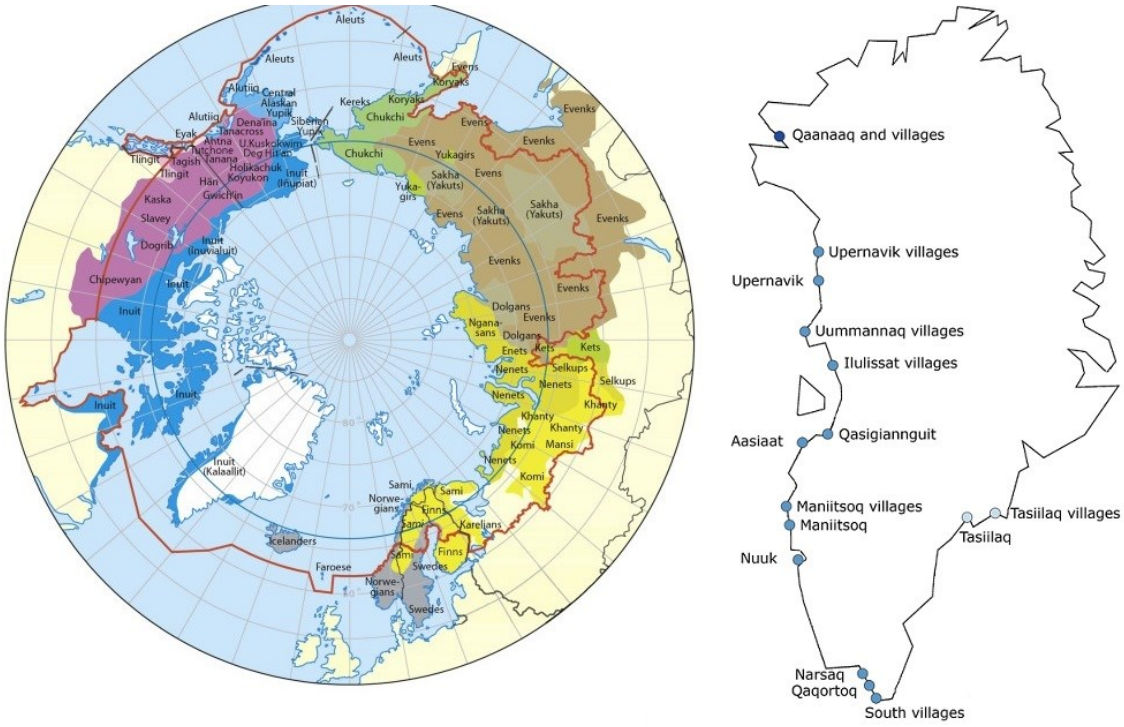

Fig 1. The map of circumpolar indigenous peoples (T.K. Yong *et al.* (2012)) and the villages of Greenland (I. Moltke *et al.* (2015))

The Peopling of Arctic Canada and Greenland: the anthropological story

From Paleo- to Neo-Eskimos

Ancient Arctic peoples in North America, ancestors of present-day Inuit, are commonly believed to comprise two distinct cultures: Paleo-Eskimos and Neo-Eskimos. Early Paleo-Eskimos have peopled Arctic Canada and Greenland in the era of 3,000 to 800 BCE, the source of their food was essentially caribou and musk ox[8], and they had a culture which was very similar to the one of Siberian Neolithic peoples. Early Paleo-Eskimos came from two consecutive waves of migrations:

1) the pre-Dorset culture and 2) the Independence I-Saqqaq culture in Arctic Canada and Greenland, including a small culture of Groswater in Labrador and Nunavik[9]. During the cold period of 800 BCE, the early Paleo-Eskimos developed into late Paleo-Eskimo, namely Independence II-Dorset culture in the eastern Arctic, and it was at this point that they started to mostly consume the products of sea mammals[10]. The Dorset culture was further divided into four phases[11], although whether they represented genetic continuity or not is controversial, and eventually ended in around 1,350 CE, replaced by the Neo-Eskimo (Thule) culture, which originated from Bering Strait region at the earliest 200 BCE[12], during the terminal phase of the Dorset culture (**Fig 2**).

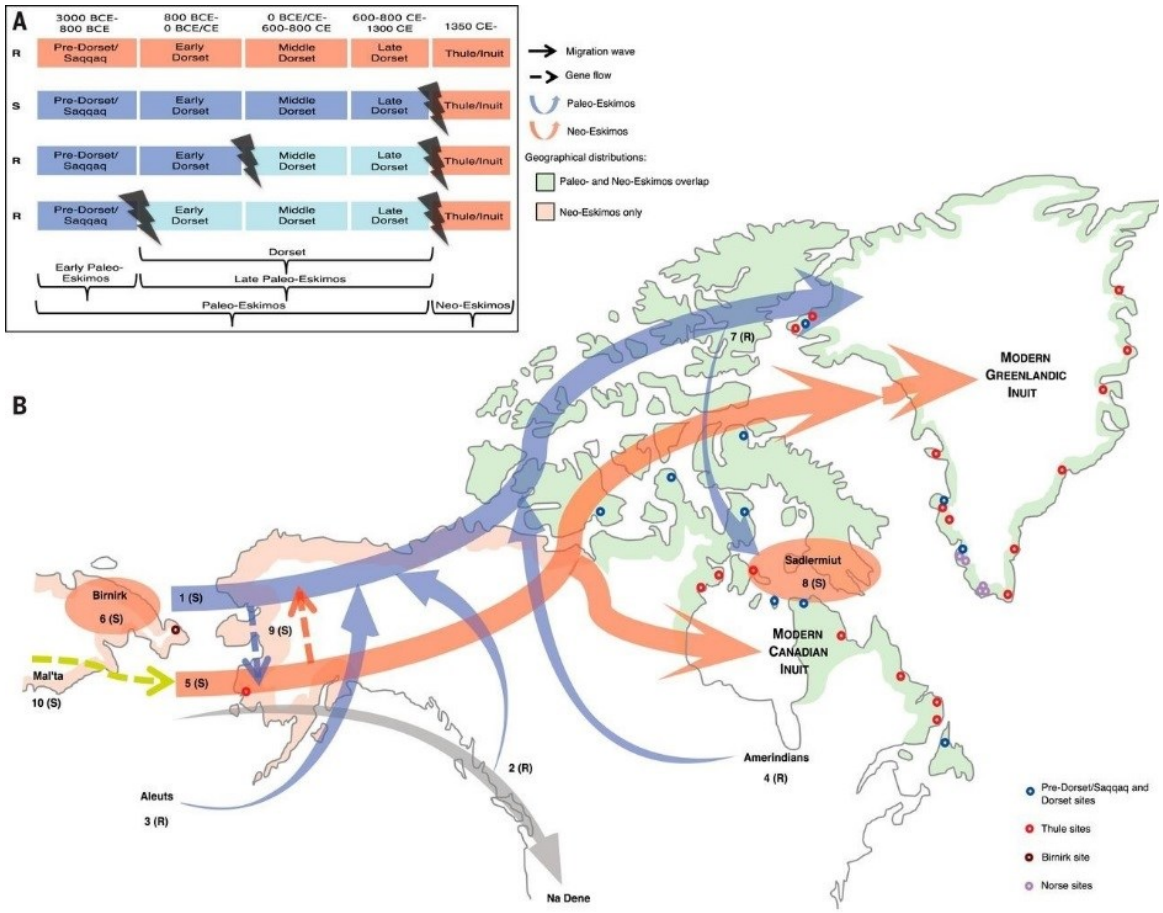

Fig 2. The genetic origins of Paleo-Eskimos and Neo-Eskimos (Raghavan *et al.* (2014))

Three cultures have led to the Thule culture: starting with the Old Bering Sea stage from 200 BCE to 500 CE; followed by Punuk and Birnirk stages from 800 to 1,400 CE, which distributed along the major Strait islands to shores of the Chukchi Peninsula, and along coastal northern and western Alaska, respectively[13]. Starting at 1,100 CE, the classic Thule culture began their movement eastward, during which they gradually changed from mainly whale hunting to hunting of walrus and seal, making them the ancestors of the present day Alaskan Yupik/Inupiat and Canadian/Greenland Inuit[10]. There was a coexisting wave of Norse Viking settlement, which came around 985 CE, and who mainly occupied the Southwest Greenland and part of Newfoundland and Labrador, introducing an ancient admixture to the Inuit peoples in those regions, combined with a recent Danish admixture, resulting an approximate 25% European ancestry in Greenland Inuit[14].

The Population History of Greenlandic Inuit

The Greenlandic population is a historically small one, as it currently consists of approximately only 57,000 individuals. The colonization of Greenland followed its route with the peopling of North American Arctic: Paleo-Eskimos entered the Northwest Greenland via Canadian High Arctic Island at 2,500 BCE[15], which was eventually replaced by the Thule people in the 12th century. This hypothesis was further supported by the genome of a 4,000-year-old Saqqaq individual retrieved in Qeqertasussuk, Greenland; this sequencing enabled a link to be made between the Old World Arctic population and the extinct Greenlandic inhabitants[16]. Past genetic studies have reported evidence that highlighted how modern-day Inuit in Greenland are direct descendants of the first Inuit pioneers of the Thule culture[9]. MtDNA analyses also led to conclude that Thule Inuit may have encountered and interbred with concurrent Dorset individuals in Canada

and on the east coast of Greenland[9; 15], thus suggesting that current Greenland Inuit are the descendants of both Thule and Dorset cultures[17]. This hypothesis is however disputed by other groups[14]. Afterwards, Norse Vikings settled with Dorset and early Thule people in Greenland from 985 until 1,450 CE[18; 19], Thule Inuit started their settlement in the West and Southeast of Greenland from 14th century. Analyses of the Greenlandic migration events supports the hypothesis proposing a single wave of Inughuit Polar Eskimos to have moved into Northwest Greenland in the 17th century[20], and then moved southward along the west coast[14]. In 1721, Greenland became a Danish colony until 1979, introducing a large number of European settlements on the island for the second time in history[21].

Inuit in Canadian High Arctic

Current Canadian Inuit live mainly in the subarctic region of Nunavut and Nunavik regions of Quebec, a small number of them also live in the Nunatsiavut and NunatuKavut regions of Labrador and Arctic Ocean of the Northwest Territories. The Nunavik region is 443,685 km^2 in size and is composed of 14 Inuit communities, with a total population of 12,090. The occupying of Nunavik started 4,500 BP by Paleo -Eskimos, at the east coast of Hudson Bay and Ungava Bay, until 2,500 BP because of the climate changes. During which the Inuit population had gone through several declines and a resurgence from 3,000 to 2,800 BP[22], until they were eventually replaced by the highly cold-adapted Dorset people, who reached Nunavik around 500 BCE. The Dorset people were subsequently replaced by Thule around 1,300 CE, during the Medieval Warm Period that had a devastating impact on their ways of hunting on sea ice. The contact of Nunavik Thule with Europeans, contrary to that of Greenland Inuit, remained near non-existent until the 19th century. Modern day Nunavik inhabitants are mostly descendants from the Thule culture, while a small

number of late pre-contact Innu sites are also found[9]. Innu were a nomadic people who likely had spent the colder months hunting caribou inland before visiting coastal areas in the spring and summer to hunt marine animals, and with ancestors likely to be Paleo-Indians and developed into a new culture known as the Maritime Archaic people.

The peopling of Nunavut from the starting of pre-Dorset culture to the Thule occupying was similar to that of Nunavik, but with an earlier inhabitation: when the glaciers in Nunavut melted approximately 10,000 BP, Paleo-Indians from the first wave of migration reached this area and moved northward following the migration of caribou to the tree line. At around 500 BP, with the advancing of Little Ice Age, Nunavut Inuit abandoned most regions of the High Arctic and had their first encounter with Europeans[12].

The Links between the Past and the Present

There have been some debates as to whether the Paleo-Eskimos were an offshoot population to the one that gave rise to the Native American population. Ancient mtDNA studies suggested that the first wave of migrations, which colonized the New World Arctic, originated neither from the Native American populations started 14,350 BP nor from the populations that gave rise to the later Neo-Eskimo expansion from Alaska around 1,000 BP[23]. Several studies also aimed to establish if there was genetic continuity between Paleo-Eskimo and today's Inuit. Raghavan *et al.* pointed out that all Paleo-Eskimos appear to be a continuum of a single-source population[9], although gene flow between the Dorset and the Thule Inuit could not be excluded. A recent mtDNA study reported Alaskan North Slope being the starting point for both Paleo- and Neo- Eskimo migrations further east-ward into Canada and Greenland, and thus there were some degree of genetic continuity between Paleo-Eskimos and contemporary Inuit peoples[24]. An independent mtDNA

analysis showed that Neo-Eskimos were mainly of the Arctic Canada origin, whereas the direct ancestors of Paleo-Eskimos were primarily drawn from Chukotka[25]. It has also been suggested by Dulik *et al.*[26] and Olofsson *et al.*[17] that there has been past gene flows between Dorset and Thule Inuit, at least in Greenland, which bear traces of both Thule and Dorset descent. Therefore, the current populations of the North Circumpolar Region share a common origin with Native American populations.

The Genetic History of New/Old World Arctic Peoples: evidence from DNA

Reconstructing Greenlandic genetic history

The genetic structure of Greenland Inuit has recently been extensively described, through large scale studies using both high density SNP genotyping array and high-throughput sequencing methods[14; 27]. The Greenlandic Inuit population came from small founder populations and they encountered a series of bottlenecks in history, consequently they would have had relatively small effective size and are expected to harbor less genetic variations; the latter has been proved by a number of subsequent studies. Pereira *et al.* focused on X chromosomal markers of 477 males from Greenland (N=312) and Denmark (N=165)[27] reported that Greenland Inuit have a reduced diversity and a higher level of LD by comparison to the Danish and Asian populations. Their study showed a significant difference between Danish with all Greenlandic sub-populations, especially the Ammassalik. Ammassalik and Nanortalik (**Fig 1**) presented the most significant genetic distances to the reference populations compared with the other Greenlandic populations. Moltke *et al.*[14] further advanced the study of Greenlandic LD by reporting that in the ancestral Inuit population, their LD was markedly higher than the present-day Greenlandic population, where the

recent European admixture has reduced the LD of the Greenlanders. The high level of genetic distinction in Greenlandic populations was also observed in a study from Tomas *et al*. where the pairwise F_{ST} between multiple populations (including Somalis and Danes) showed a clear displacement of the Greenlandic population versus the others[28]. Identical by descent (IBD) analyses suggested that approximately 84% of the genetic differentiation inside the Greenlandic sub-populations may be due to genetic drift, and a different degree of non-Inuit gene flow within the different subgroups[14]. Admixture analysis also revealed that Greenlanders present a low level of the European/Danish ancestry and also a low level of Asian ancestry.

Moltke *et al.* conducted a more extensive modelling in the Greenland Inuit by genotyping 4,674 Greenlandic individuals with a MetaboChip that contained 196,224 SNPs[14]. Their first finding was consistent with the previous findings of European ancestry admixture as they observed that there has been gene flow from Europeans into most of Greenland. Over 80% of Greenlanders have European ancestry, but the distribution is not even across Greenland (individuals from Ammassalik, the South Villages and Qaanaaq in the Northwest Thule region have less European ancestry or only Inuit ancestry). The findings could be interpreted with a scenario where Greenland was settled by Inuit who entered North Greenland and migrated to South Greenland along the west coast and later to East Greenland. Nonetheless the Greenlandic gene flow is reported to be sex-biased, which was correlated with an earlier finding of Bosch *et al*.[21]. Although as much as 50% of Greenlanders have European male ancestry, only 1% of the female ancestry were European. There is also population differentiation even within the Greenlandic Inuit population – as the observed F_{ST} between Upernavik villages and South villages is smaller than the estimated between Upernavik villages or other locations, thus suggesting that Upernavik villages are genetically closer to South villages (**Fig 1**). Ammassalik from Southeastern Greenland are genetically furthest

away from Qaanaaq in Northwestern Greenland (F_{ST}=0.04), as they are closest to the villages in South Greenland (F_{ST}=0.02). This slight decay of genetic variation following a north-west-south-east gradient of Greenland further supports the aforementioned hypothesis proposing a single wave migration along the coast.

Additionally, the genetic-drift tree topology and D statistics relating Greenlandic sub-groups also inferred the existence of a single coastal route migration, in which each location sequentially split off along the coastline. On the other hand, the D statistic test also suggested that there had been only limited amount of gene-flow from the Dorset to the modern east Greenlanders, when compared with the ancient Saqqaq genome[16]. The last question that Moltke *et al.* addressed in their study was if there were Norse Viking admixture in the Greenlandic genome by analyzing the length of ancestry tracts, as ancient admixture tends to have shorter tracts. Unfortunately, they reported no significant differences when comparing the tract-length distributions of Northwest villages and South Villages, where little recent European admixture was found and large settlements of Norse Viking once took place.

Beringia peoples lived in cold climate conditions and their diet was rich in fat and protein from fish and marine mammals. It is likely that all Native Americans and related groups like Inuit descent from populations originally have adapted to such conditions, since the migration route into the Americas for all of these groups was through Beringia. Fumagalli *et al*.[29] thus followed up the previous population structure study of Greenland[14] by analyzing the environmental adaptation of Greenlandic Inuit. Through analyzing the same SNP-chip dataset[14; 30], they selected 191 Greenlandic individuals who have less than 5% estimated European ancestry and applied population branch statistic (PBS) for the detection of positive selection signals. As a result, they found the strongest signal of selection on chromosome 11, which comprises multiple genes

(*TMEM258, MYRF* and *FADS1, FADS2* and *FADS3*). They also reported other loci that appeared to be associated with selection signal, e.g. on chromosome 1 (encompassing *WARS2* and *TBX15*) and chromosome 17 (*FN3KRP*). To supplement the SNP-chip data, they generated WES data on 18 Greenlandic Inuit and observed two additional genes, *DSP* and *ANGPTL6*, representing the strongest signals of selection. Although the selection signal for *FADS* cluster was reduced using the WES dataset (PBS=1.145, *FADS2*), the report by Fumagalli *et al*. provided additional evidences that highlighted the selected alleles in genes known to modulate fatty acid composition and affect the regulation of growth hormone, which would have been critical for the adaptations to a diet rich in PUFAs.

The Genetic Structures and Selection Signals in Siberians

Studies of the Old World peoples can also be informative in analyzing contemporary Inuit populations who have adapted to the cold climate before they evolved through their own courses, possibly due to the result from the bottleneck drift that New World peoples experienced during their period of migration. Some Siberians appear to have different cardio-metabolic disease risks than other circumpolar groups[31]; and other studies observe lower serum lipid levels in indigenous Siberian peoples[32].

Several mtDNA studies examining Siberians preceded the era of genome sequencing and these provided the first genetic evidence supporting archeological findings that led to believe that the High Arctic populations originated in Siberia. These mtDNA studies found the latest inhabitants of Beringia to be genetically linked with the Chukchi-, Eskimo-Aleut- [33], and Na-Dene-speaking Indians as well as Paleosiberian-speaking Yukaghir in the high arctic since early Postglacial[34]. These studies also established the founding population of Eskimo-Aleut to have

originated in Beringia/Southwestern Alaskan during the early postglacial period[35], and today's Greenlandic individuals to essentially be the descendants of Alaskan Neo-Eskimos[36].

Studies that included Siberians for the identification of selection signals showed that the most strongly selected signals across worldwide populations are associated with pigmentation, autoimmune diseases and genes related to UV radiation, infection, immunity, cancer and metabolic adaptations[37; 38]. In 2014, Cardona *et al.* conducted a study describing the population structures of indigenous Siberian peoples and their selection signatures[39]. They used a SNP-chip that contains 730,525 SNPs in 200 individuals from 10 Siberian groups. The initial assessment using principal component analysis (PCA) and ADMIXTURE showed different ancestry profiles that enabled distinctions to be made across native Siberian populations. Among those distinctions it appears that Chukchi, Eskimo and Koryaks were predominately of Inuit ancestry; also Yakuts, Evens and Evenks came mainly derived from another ancestry; and finally Altai, Teleuts, Shors and Buryats all shared a third. The clustering of different predominant ancestries into different groups was fitting with their respective geographical locations. Subsequently, two haplotype-based tests (iHS and XP-EHH) were used to examine the positive selection, and seven genes (*THADA, ITPR3, GNGT1, PRKG1, RELN, CPT1A* and *LRP5*) were observed to be positively selected; furthermore a locus on chromosome 11 containing the last two genes was delineated and shown to be the most significant. Interestingly, this region was close to the region that had been found to be positively selected in the earlier Greenlandic Inuit study by Fumagalli *et al.*[29]. Cardona *et al.* demonstrated that this region contained the strongest signal in the Northeastern Siberian gene pool[39], the ancient ancestors of today's Greenland and Canadian Inuit.

A followed-up study by Clemente *et al*.[40] based on the aforementioned chromosome 11 region generated whole genome sequencing (WGS) data from 25 individuals from Chukchi,

Eskimo and Koryaks[39]. After a careful variant filtration, 148 SNPs with both a >0.5 derived allele frequency in subjects of the study and a MAF<0.01 in published populations were retained. From which they reported three variants with possible functional consequences to the encoded proteins, which located in two genes, *CPT1A* and *IGHMBP2,* within the previously found chr11 region[39]. iHS test focused on this region with the inclusion of other populations mapped chr11:66-69 Mb region to be the significantly selected locus only in Northeast Siberians. Tajima's D scans further narrowed down a region of 68.2-68.6 Mb, which contains genes *LRP5, PPP6R3*, *GAL, MTL5* and *CPT1A*. This study eventually established that a variant in *CPT1A* (c.1436C>T) drove this selective sweep, and this will be further discussed in the following section.

Discovering Selection Signals in FADS and CPT1A in Arctic Populations

Fumagalli *et al.* suggested that in Greenlandic Inuit, the strongest selection signals are in the *FADS* gene cluster[29]. The family of *FADS* genes are known to be associated with long chain polyunsaturated fatty acids (PUFA) metabolism, type 2 diabetics (T2D) and coronary artery disease[41-43]. Using the inferred demographic model by Approximate Bayesian Computation (ABC), the divergence time between Han Chinese and Greenland Inuit was estimated to have occurred 23,250 years ago with an effective population size of 1,550. The chromosome 11 region was thus established to be a strong outlier and the selection starting time of the most significant SNP (rs74771917) from the *FADS* cluster was occurred approximately 19,751 BP (selection coefficient s=0.0313), when the common ancestors of Inuit and Native Americans still lived around Beringia, and before the first migration wave to the New World started.

A positive selection of the *FADS* cluster has been reported in many other populations because of its involvement in n-3/n-6 PUFA metabolism, a key component for the diet evolvement

of Homo sapiens[44], (e.g. in Africans[45; 46] and Alaskan Eskimos[43]). A very recent study reported a 22bp insertion-deletion (rs66698963) in *FADS2* that was positively selected in Asian populations, more particularly in South Asians[47]. *FADS2* was positively selected in both vegetarians, such as South Asians; and meat eaters, such as Greenland Inuit. Interestingly, the ancestral allele of rs174570 in *FADS2,* which was positively selected in Greenland Inuit, also exhibited positive signals in South Asians. The differences being that the derived allele was under adaptation in Greenland Inuit, while in South Asians, the ancestral allele was better adapted. It is tempting to suggest that the ancestral haplotype of *FADS2* variants has been selected towards dairy-vegetable diet while the derived haplotype has contributed to the adaptation to a marine-fat-rich diet.

Another interesting gene is *CPT1A* for which a missense mutation (c.1436C>T, p.P479L) was reported in multiple studies; most likely because of its high frequency (possibly due to natural selection) across Arctic populations (0.77 in Nunavut and 0.44 in the Northwest Territories)[48; 49]. This Arctic-specific variant was initially thought to be associated with CPT1 deficiency, including hypoketotic hypoglycemia[50] and sudden infant death syndrome[51]. A functional characterization of the *CPT1A* c.1436C>T mutation revealed that this mutation decreases the process of fatty-acid oxidation and ketogenesis under normal condition. However under conditions of high malonyl-CoA concentration, the activity of CPT1A c.1436C>T mutation was much greater than what could be measured in the cells with the wildtype allele[52]; which could consequently have a protective role against the overproduction of ketone bodies[53].

In order to study the origin of this mutation at the evolutionary level, Clemente *et al.*[40] performed LD analysis focused on chr11:68-69Mb region that had shown strong selection in their previous study[39]. They found that *CPT1A* c.1436C>T mutation seem to be the driving selection (MAF=0.68) in Northern Siberians as a selective sweep[54], as it forms two distinct LD blocks only

in Northern Siberians while the pattern is not significant in East Asians and completely absent in Europeans. The c.1436C>T variant was thought to be absent from the genomes of most individuals worldwide, and to be only presented in Arctic populations. The inferred demographic model of Northeastern Siberians yielded an effective population size of 3,000-5,000; under the assumption that c.1436C>T variant is the result of directional selection from a *de novo* mutation, they estimated the selection time of this variant to be 3-ka old (selection coefficient s=0.14). The age of selection time of a 58kb haplotype around the c.1436C>T variant is around 6.7-ka, supporting the hypothesis of a strong selection from a selective sweep, possibly due to a recent and rapid adaptation against the cold climate and dietary specialities. To concur with the estimation of the age of this mutation, they examined ancient samples for its surrounding haplotype (comprised of 27 variants in high derived allele frequency in Northeast Siberians). They found that this derived haplotype was absent from the Neandertal and Denisovan genomes, regardless of the fact that *CTP1A* was also found to be among the lipid-catabolism-associated genes enriched with Neandertal-like sites[55]. It was also absent from the ancient human genomes more than 10-ka old[56], but starting to appear in the 4-ka old ancient Saqqaq genome[16], then also in Paleo-Eskimo/Neo-Eskimo samples[9]. Interestingly, this haplotype was partially presented in Aleutians but completely absent from Na-Dene speaking Athapaskans, perfectly consistent with a previous conclusion that Eskimoan- and Athapaskan-speaking populations are genetically distinct and came from two population expansions that occurred after the initial wave of migration[26]. It was therefore postulated that the c.1436C>T variant (and its surrounding haplotype) could have occurred in the Old World Arctic (ancestors to the Northeastern Siberians) and only retained by the selection pressures of cold climate, before it was introduced to the New World Arctic, starting with the first Paleo-Eskimo wave that populated the Canadian High Arctic and Greenland.

Founder Mutations and Disease Risks in Inuit Populations

The recent changes of Inuit lifestyle, such as drinking and smoking, calorie-dense foods and less hunting activities, were believed to have increased their risk of cardiovascular diseases (CVD), including atherosclerosis and hypertension, compared to western populations. Parallel studies comparing Inuit living in Greenland and Denmark[57] also supported this belief. Considering these observations it can be hypothesized that variants which neutrally accumulated in the Inuit populations may be starting to show deleterious effects under the new selective pressure. On the other hand, because of the high degree of genetic differentiation following the separation from East Asian or Beringian founder populations[58], causal variants that are very rare in other populations may segregate at a distinctly high frequency in the Inuit population, making it a suitable model for discovering disease susceptible variants.

Metabolic Traits and Diseases

The best example of this scenario would be the p.R684X variant in *TBC1D4*, which was discovered through a genotyping and WES study examining 2,575 Greenland Inuit, and found to be associated with a higher 2h plasma glucose levels[30]. The discovery phase of the GWAS found the highest association to be with an intronic variant (rs7330796) of *TBC1D4* and subsequent followed-up WES and further genotyping in general Greenlandic cohort and T2D cohort revealed the causative variant p.R684X in *TBC1D4*. This nonsense mutation was shown to be strongly associated with mean 2h plasma glucose levels (mainly in homozygous carriers), and with an increased risk of T2D, which elevated circulating glucose and insulin levels after the glucose load. The p.R684X

variant has a large population impact, with an estimated frequency of 23% in the Inuit cohort, possibly due to a weak selection or genetic drift; while it is absent from Europeans and only at a frequency of 0.5% in 1KGP Japanese population.

More studies have focused on the association of Inuit specific lipid-metabolic profiles with their genetic background. In the selection study of Greenland Inuit carried out by Fumagalli *et al.,* a strong association was observed for a positively selected *FADS* SNP with body weight ($p=1.1\times10^{-6}$; rs7115739) and height (p=0.00012; rs7115739), FADS affects the fatty acid desaturases by compensating for a high dietary intake of eicosapentaenoic acid[29]. In Greenland Inuit, a rare haplotype in the gene encodes adipose triglyceride lipase (ATGL) was found to be associated with increased plasma triglyceride concentrations[59]. The allele distributions in several *ALDH* genes, particularly *ALDH1B1* and *ADH1C*, shown to be associated with drinking patterns, also differed significantly between Inuit and Danes[60]. The Inuit also harbor distinct variant distributions in other genes (*CYP1A1, CYP1B1* and *COMT*) and this suggests their susceptibility to environmental contaminants to be increased[61], the same variants may also contribute to their predisposition to diabetes[62]. Additionally, a study of 553 Nunavik Inuit, focused on 35 variants across 20 lipid-metabolism related genes pinpointed 7 SNPs as significantly associated with plasma lipid parameters[63]. And genes such as *DOCK1, PTPRE* and *FABP4,* were highly associated with obesity-related traits and n-3 PUFA interactions in Yup'ik people[64]; SNPs in *LPL, TNFRSF10B*, and *APOJ* also had strong statistical evidence of functional effects in the fatty acid distributions[65]. *CDKAL1* and *HHEX* polymorphisms were also reported to be associated with glucose homeostasis in Alaska Native population with a low prevalence of T2D[66].

Disrupted profiles of the metabolic genes sometimes could lead to more severe conditions. In a large cohort that regrouped Inuit from Alaska, Canada and Greenland[67], the *LDLR* p.G116S

variant was observed to be strongly associated with dyslipidemia; a condition found in approximately 10% of Inuit and with one of the largest effect sizes (3.02-fold increased risk of hypercholesterolemia). In another Nunavik Inuit study examining glycogen storage disease type III (GSD-III), a founder mutation, *AGL* c.4456delT[68], was reported to cause the disease in individuals who are homozygous carriers. It should however be noted that the same homozygous variant was also observed in a cohort of GSD-III cases of North African Jewish heritage, albeit it is absent in other present-day populations. The reason why this old variant was somehow preserved only in these two very genetically distinct populations is unclear. Finally, a variant c.273_274delAG located in the gene encodes *Sucrase-Isomaltase* (*SI*) was also found common in the Canadian Inuit population (MAF=17.2%) with homozygous carriers develop congenital sucrase-isomaltase deficiency[69].

Infection and Allergy

The risk of infectious diseases is another health concern for Inuit populations. For instance, a high risk of HPV infection was observed among Nunavik Inuit women[70]; HLA-G polymorphisms were also suggested to affect the host immune recognition and this is supported by the fact that Nunavik Inuit women have a distinctive HLA allele frequencies by comparison to other populations[71].

A common Inuit variant *IL4RA* Ala57Thr, while extremely rare in other populations (MAF≤0.6%), was found associated with the risk of atopy in Inuit living in Denmark, but not in Greenlandic Inuit[72]. This finding was not an isolated case as in Inuit populations of Greenland and Denmark the same asthma risk allele expresses opposite magnitude and direction of genetic associations[73]. The Inuit population separation time between Greenland and Denmark is however

too short for the effect of a genetic drift to have taken place, suggesting that environmental changes also play a big part in some of the disease pathogenesis observed in Inuit.

Cancer

Inuit populations have recently observed to present with an elevated risk of cancer. A founder mutation c.2002A>G in *PMS2* was observed in a Nunavik Inuit family with constitutive mismatch repair deficiency (CMMRD) syndrome, which leads carriers to eventually develop cancer[74]. The intronic variant creates a *de novo* 5' splicing site for intron 11 that leads to a frameshift. Nine additional Inuit cancer patients were subsequently discovered to be homozygous carriers of the *PMS2* c.2002A>G; this variant is therefore likely to act in a recessive inheritance manner. The variant has an estimated frequency of 0.06 in the Nunavik Inuit populations and is considered to be one of the most common cancer-associated risk allele.

Breast cancer risk is high amongst Greenland Inuit women and reports have highlighted a founder mutation, *BRCA1* p.C39G, that is observed in 33% of ovarian cancer patients and 10% breast cancer patients, the overall frequency of this variant is 1.6%[75; 76]. However the frequency of this variant differs geographically as it is 0.6% on the West Coast to 9.7% on the East Coast, a difference that likely resulted from bottleneck drift during the migration. Other polymorphisms in *CYP1A1* and *CYP17A1* can also increase the risk of breast cancer among Inuit women and the risk increases with higher serum levels of Perfluorooctane Sulfonate (PFOS) and Perfluorooctanoic acid (PFOA), which were consistently shown as risk factors of breast cancer as well[77].

By opposition, Inuit men from Greenland have lower risk of prostate cancer. Several *SRD5A2* SNPs and *androgen receptor* repeats were shown to differ significantly between

Greenland Inuit and Swedish, these differences lead to a lower SRD5A2 enzymatic activity that regulates the turnover of testosterone to 5 alpha-dihydrotestosterone[78].

Liver Diseases and Others

An *ATP8B1* c.1660 G>A variant was reported to cause cholestasis familiaris groenlandica, a liver disease common in Greenlandic Inuit[79], the variant causes the disease in a recessively inherited manner and it is found in an heterozygote state in 12% of the population[80]. The observation of a *BLVRA* c.214C>A nonsense mutation in unrelated Inuit women with biliary obstruction during cholestasis suggest it might be a founder mutation[81]. A frameshifting mutation *ATP13A2* c.2473C>AA was also found in a Greenland Inuit family affected with Kufor-Rakeb syndrome (KRS)[82].

Inuit hearing impairment, a highly prevalent disorder in this arctic population, was reported to be caused by a *GJB2* c.35delG mutation, the variant frequency in overall population is found vary between 0.5% in the West Coast to 2.3% in the East Coast[83].

Cerebrovascular Diseases

Whether or not related to the high risk of CVD that the Inuit populations recently acquired, there is also an elevated risk of cerebrovascular diseases in Inuit[5], the genetic underlying of which may be combined with heavy alcohol consumption and hypertension, known to increase the CVD risk. A report of CVD prevalence among Nunavik Inuit indicated that 2.5% of adults were diagnosed with a cerebrovascular disease[84]. In Alaska, cerebrovascular disease is the 5th leading cause of death (4.5%) and the mortality rate is 1.5 times higher than the rate observed in the European

descendants[85]. Lindgaard *et al.* reported that Inuit IA patients have a much higher rate of familial history of subarachnoid hemorrhage (SAH) (23.1%) and IA (9.6%) compared to Danish IA patients (4.3% and 1.6%)[86]. Moreover the relative risk of SAH of Greenlandic Inuit is 4.4 fold higher than the one observed for the neighboring Dane population[87]. The prevalence of IA in Denmark is similar to that of Sweden and Finland[88], individuals from the latter also deemed to present a high risk of IA[89]. Finally it is noteworthy that Inuit IA cases also tend to harbor multiple IAs, the percentage of such cases is two-fold higher than Danes; fitting with a genetic risk factor in Inuit since multiple IAs is a marker for the genetic form of IA. Despite this evidence, the underlying genetic cause of IA in Inuit and the contribution played of founder mutations nonetheless remain to be further elucidated.

Conclusion

Through several waves of migration, ancient groups of humans began populating the Americas nearly 16 millenniums ago. The arctic population is an interesting one as it provides us with an extended picture of how ancient human occupied the New World, it also sheds light on how today's global populations came into being. Through examinations of their genomes which were adapted to the harsh arctic environment and affected by bottleneck drifts, the arctic indigenous peoples like Inuit might lead us to discover not only the genetics that explain specific human traits but also the underlying mechanism of certain disease etiology. Canadian Inuit are one of the most important residents of the Arctic, but their genetic landscape remained a largely uncharted territory. Further studies focusing on this unique population could therefore efficiently link the studies of Siberians and Greenlandic Inuit and advance our knowledge of the global population genomic profiles.

1.3: THE GENETIC ETIOLOGY OF INTRACRANIAL ANEURYSM

Cerebrovascular diseases, stroke and subarachnoid hemorrhage

Cerebrovascular diseases include a broad spectrum of disorders affecting the blood supply to the brain. Stroke, carotid stenosis, IA, artery malformations, including arteriovenous malformations (AVM) and moyamoya disease, all belong to this heterogeneous disease group.

Stroke is the second most common cause of death after coronary artery diseases in the western contries[90], while in China, it is the most frequent cause of death, with incidence more than five folds that of myocardial infarction. It can be categorized into ischemic (blockage) and hemorrhagic (bleeding), with 87% being ischemic[91]. Ischemic strokes are commonly due to thrombosis or embolism while hemorrhagic strokes, which are subdivided to intracerebral and subarachnoid, are usually caused by an arteriovenous malformation or the rupture of an aneurysmal lesion.

SAH is bleeding in the subarachnoid space, an area between the pia mater of the brain and the arachnoid membrane. SAH comprises approximately 5% of all strokes, with an incident rate of 6-7 per 10,000 person/year[92] and a fatality rate around 50%[93]. The prevalence of SAH is higher in Finnish and Japanese[94; 95], and surprisingly this is not due to their prevalence of IA[96].

The risk and etiology of intracranial aneurysm

The risk of IA rupture

An IA is a dilatation of part of the intracranial artery and its shape varies from a localized sac-like pouch to a long enlargement of the vessel diameter[97]; these two forms are referred to as saccular or berry aneurysm and fusiform aneurysm, respectively. The worldwide incidence of intracranial aneurysms has been estimated to be approximately 2-3%[98], most IAs are saccular (SIA), which typically occur at the arterial branching sites on the Circle of Willis. Common sites of SIA include the anterior communicating artery, the bifurcation of posterior communicating artery and internal carotid artery, the middle cerebral artery, and the basilar artery bifurcation[99]. The most common clinical symptoms of aneurysmal SAH are a sudden onset of severe headache with stiff neck, vomiting, and photophobia[100]. Unruptured IA (UIA) is usually asymptomatic, however retrospective studies have estimated the annual risk of rupture among IA patients to be approximately 1-3%[101].

Although ruptured IAs are the cause of 85% patients who had subarachnoid haemorrhage, it has been reported that 50% to 80% of all IAs do not rupture during the course of a person's lifetime. A recent examination of the prevalence of UIA estimated it to be around 3.2%[96], and can be as high as 7% in China. The prevalence ratio of UIA is significantly higher in women than men (1.61) and in individuals over 50 years versus less (2.2)[96]. 1.8% of the people with European ancestry have SIA[102], the incidence usually varies among different populations, with the highest in Finnish and Japanese. However, recent studies suggested those statistics to be disputable[96; 103].

The risk of SIA is also higher in some connective-tissue related genetic disorders. Autosomal dominant polycystic kidney disease (ADPKD) is the most common disease associated with SIA, with a relative risk of 6.9 (95% CI 3.5-14). Approximately 5% to 40% of patients with ADPKD have IA, and 10% to 30% of patients have multiple IA[96]. Other associated disorders

include neurofibromatosis type I[104], Ehlers-Danlos syndrome[105], Marfan syndrome[106], hereditary hemorrhagic telangiectasia[107] and endocrine neoplasia type I[108].

The risk of UIA to rupture increases with the aneurysm size, location (particularly in the posterior circulation), irregular shape, and history of previous aneurysmal SAH[109]. Other studies suggested smoking and female sex increase the risk of IA rupture, size of the IA lesion, on the other hand, is not a significant risk[110]. Controversies always follow the size of aneurysm: small aneurysms have low annual risk of rupture[111]; however, 85-90% of the ruptured aneurysms are small[112]. A more recent study reported the risk of SAH depends predominantly on smoking, sex and blood pressure[113]. Excessive alcohol consumption also increases the risk of SAH[114].

The exact etiology of IA is still unclear, albeit several hypotheses highlight the contribution of maladaptive vascular remodeling[115] triggered by hemodynamic stress[116] and inflammatory response[117], such chain of events would ultimately damage the blood vessel wall, which is the key process in the pathophysiology of IA.

Intracranial arteries and IA pathology

Intracranial arteries have two major differences compared to other arteries: 1) there is no elastic lamina and the adventitia is thinner in the vascular walls[118]; 2) there are gaps in the muscular media layer in the bifurcations[119]. These special characteristics may help to explain the site of the aneurysm formation.

IAs have different arterial wall types that characterized by their specific histopathology: 1) perturbations of homeostasis in the cerebrovascular wall, characterized by a lack of internal elastic lamina in the IA wall[120], which occurs early in the course of the disease; 2) decrease in tissue

cellularity along with changes in the extracellular matrix composition, with the number of smooth muscle cells (SMC) decreasing and wall thinning[121], as well as apoptosis and necrosis[122]; and 3) hyperplasia of the tunica intima and disarray of vascular smooth muscle cells; followed by an increased risk of rupture in thicker walls with myointimal hyperplasia and thrombus[123].

Changes in composition of the vessel wall are also observed in IA tissues as the levels of collagen, laminin and several extracellular matrix components are decreased[124]. Type I collagen and fibronectin appear dispersed in IA wall while in normal artery the expression of type I collagen is restricted to the adventitia and fibronectin to the media[125]; type III and IV collagen and laminin also appear to have decreased expression in IA wall[124]. Atherosclerosis is another common pathological feature in IAs and it is noteworthy that the progression of atherosclerosis correlates positively with aneurysmal growth[126].

Hemodynamic stress

IAs tend to occur at regions of sharp vessel curvature, and also occur more frequently at certain bifurcations, such as the anterior communicating artery and middle cerebral artery. Because those regions experience high wall shear stress and other hemodynamic forces, it is assumed that local hemodynamic environment is critical to IA formation[127]. Animal studies have demonstrated that elevated hemodynamic stress could trigger aneurysm-type remodeling at a bifurcation site[128] and at the basilar terminus[129].

Endothelial cells can undergo significant changes, both in function and morphology under hemodynamic stress[130]. In response to shear stress, endothelial cells may elongate, change alignment while developing actin stress fibers and may change cell migration[131]. Hemodynamic

stress may also relate with changes in molecular signaling, such as mediating pathway of inflammation, within the endothelium and amplifying chronic inflammation[132].

Studies focusing on correlating hemodynamic stress and IA formation established the following facts: 1) IA occurs more often in locations experiencing higher wall shear stress (WSS) and wall shear stress gradient (WSSG); 2) the susceptibility of IA reflects a threshold-type response to hemodynamics, which increases drastically at high level of WSS and WSSG[127].

Vascular remodelling

Vascular remodelling is carried out when the vascular endothelium reacts to blood flow, in which the morphology of the endothelia cells (EC) varies corresponding with blood flow patterns[133]. The reaction of EC is thought to be regulated by different factors such as calcium channel activation[134], gene expression differentiation induced by shear forces[135], mechanical events[136] and EC realignment[137].

The vascular smooth muscle cells (VSMC) in IA walls have been observed to migrate into the intima and proliferate to produce myointimal hyperplasia[126], an early event in IA formation. VSMC modulation appeared to be the most pronounced in ruptured IA: a transition of VSMC from a differentiated to a modulated phenotype and eventual loss of both in ruptured IA. Thus phenotypic modulation of VSMC in the IA wall appears to be related to the aneurysmal wall formation and to a rupture mechanism[138].

WSS is also thought to be one of the most important factor to regulate vascular structure and remodelling[139], usually through the endothelium's response. Low level of WSS tend to induce atherosclerotic lesions[140] and high level of WSS has been implicated in expansive or outward

vascular remodeling in response to a sustained flow increase[141]. High WSS contributes to the adaptive outward remodelling by 1) endothelial responses; 2) an increased stimulation of endothelial cells and smooth muscle cells proliferation; 3) an increased stimulation of matrix-degrading proteins; and 4) atherosclerotic wall remodeling[142]. Bifurcation apices and outer curve of arteries in the Circle of Willis tend to have flow impingement, which will cause high WSS and positive WSSG, thus are subject to remodelling and the predisposition to IA.

Inflammatory response

Vascular inflammation has shown to be involved in the pathogenesis of IA, although it is not yet clear if those inflammatory responses contribute directly to IA formation. Many of the pathological changes underlying IAs are related to the chronic inflammation in the vessel walls, especially in IA VSMC that has shown to undergo phenotypic modulation to a pro-inflammatory phenotype[143].

It was reported in several studies that low-density lipoproteins in VSMC are oxidized[144; 145] to produce inflammatory response in IA walls which will lead to the degeneration of vessel wall. Macrophages are probably the most important histologic indicator of the inflammatory reaction of IA walls. A recent study revealed the functions of macrophages in IA: they are critical for hemodynamically induced outward vascular remodeling and elaboration of inflammatory cytokines[146]. Macrophages contribute to the extracellular matrix remodelling by increasing the expression of matrix metalloproteinases (MMP)[147], which digests arterial wall extracellular matrix and leads to further damage via upregulation of angiogenic factors[148] and will eventually cause IA rupture[149]. Macrophages are also participants in the elaboration of the inflammatory signaling by secreting inflammatory factors implicated in the IA pathogenesis[117; 150], as macrophage-depleted mice were found to have a substantially lower risk of developing IA[146]. In addition, damage-

associated molecular patterns[151], immunoglobulins[144], lymphocytes[123], T and B cells[144; 152], cytokines[153] and growth factors[154] are all reported to be involved in the pathogenesis of IA via inflammatory infiltration. A study also found an increased number of leukocytes during the formation of IA and an induction of MMP in VSMC after mast cell degranulation[155]. Worth noting, modest hemodynamic changes might be sufficient to induce IA if the vascular wall is compromised by inflammation.

Diagnosis and treatment of IA and SAH

UIAs are often found either through screening high-risk individuals or as incidental findings of magnetic resonance imaging (MRI), particularly magnetic resonance angiogram (MRA) or computerized tomography (CT) studies for other neurological symptoms. The diagnosis of SAH is primarily by CT, CT angiography (CTA), or catheter angiography[156].

IA are usually treated either by microsurgical clipping or endovascular coiling. Both work by the occlusion of the aneurysm neck, which prevents the risk of possible re-bleeding from secondary complications. Endovascular treatment has developed in the past 20 years, which aims to reduce the operative risks of microsurgical treatment[157]. The advantage of microsurgical clipping is good durability, and can successfully treat most aneurysms[158]. The advantage of endovascular coiling is minimally invasive, decreases procedural and anesthesia time, and multiple, simultaneous treatments[159]. However, endovascular treatment requires more follow-up imaging, has increased recurrence rates, and a greater need for retreatment compared with microsurgical clipping. There are also some aneurysms with features that are not suitable for endovascular coiling. Both procedures have their benefits, although there is a study suggested that

in patients with a ruptured IA, the outcome in terms of survival free of disability at one year is significantly better with endovascular coiling[158].

The genetic susceptibility of IA

One of the major IA risk factor is genetic predisposition, which is supported by the fact that 8-9% of UIAs are familial[160] and 10% of the SAH cases are associated with a family history of SAH[161]. The prevalence ratio is also significantly higher (3.4, 95% CI 1.9-5.9) in patients with positive family history of SIA or SAH[96]. IA seems to be a disease with incomplete penetrance when occurs in families and in general with late age onset, and also appears to be multifactorial, with environmental and stochastic risk factors[162]. Throughout the years, genetic researches of IA have identified multiple risk genes and loci; however the genetic risk factors and pathogenic mechanisms for IA are far from being fully understood.

Traditional approaches: linkage and candidate gene screening

Being the earlier method of utilizing genome-wide data, linkage studies performed on the sparsely distributed markers have been applied on multiple IA families or a single large IA family. A collection of 13 major loci were found in these studies, and among those are the following: 1p34.3-36.13[163], 7q11[164] and 14q23-31[165] in North American IA families; 1p36.21-36.13[166], 19q13[167] and Xp22[166; 167] in Dutch and Finnish families, 5q22-31, 7q11, 14q22[168], 8p22[169], 17cen[170], 19q13 and Xp22[170; 171] in Japanese and Korean families; 5p15.2-14.3[172] and 13q14-21[173] in French-Canadians and 4q32[174; 175], 12p12.3[175] and 11q24-25[165] in mixed ethnicity families. Six genomic regions (1p34–36, 4q32, 7q11, 14q22-23, 19q13, and Xp22) were replicated by two or more independent

linkage studies. Three large studies focusing on multiple IA families of different ethnicities included 346 Finnish IA families[162; 176] and 542 FIA families[177].

Candidate genes involved in vascular wall and endothelia cell functions have been tested in case-control association studies of IA, especially those also discovered in previous linkage studies. Positive associations were found in SNPs located in *ELN* (rs8326, $p=1\times10^{-6}$)[178], *SERPINA3* (rs4934, $p=1\times10^{-5}$), *COL3A1* (rs1800255, $p=1\times10^{-5}$), *IL6* (rs1800796, $p=1\times10^{-5}$), *VCAN* (rs251124, $p=7\times10^{-4}$), *COL1A2* (rs42524, $p=9\times10^{-4}$), *HSPG2* (rs3767137, p=0.002)[179], *JDP2* (rs175646, p=0.004)[180], *KLK8* (rs1722561, p=0.008)[181], *TCN2* (rs1801198, p=0.002)[182], *LIMK1* (rs6460071, p=0.026)[183], *ACE* (rs4646994, p=0.03)[184] and *TNFRSF13B* (rs4985754, p=0.048)[185].

IA Gene expressions

Microarray based mRNA expression comparing IA and control tissues is a relatively unbiased approach to discover IA candidate genes and pathways. One of the first studies discovered enrichments in adherent junction, MAPK, and Notch signaling pathways[186]. A follow-up study focused on inflammatory response, discovered that histocompatibility class related genes played an important role in IA formation[187]. More studies that compared IA tissue with superficial temporal artery tissues found distinctively differentially expressed genes on certain pathways, including adhesion, proliferation and migration, extracellular matrix receptor interaction, and cell communication; as well as atherosclerosis, inflammatory response and apoptosis in SMCs[188-190].

The expression profiles of the ruptured and unruptured IA were also examined in a number of studies. A study reported that in ruptured IA, genes encode MMPs and genes involved in

apoptosis were upregulated; when compared of unruptured and ruptured IA, the matrix metalloproteinases 3 (TIMP-3) was significantly downregulated in unruptured IA, while nitric oxide synthase (iNOS) appeared to be overexpressed in ruptured ones[191]. Other studies reported inflammatory response related genes and anti-inflammatory regulators seemed to be upregulated[190; 192] in the ruptured IA. There was a similar observation with the transcription factors[193], such as Toll-like receptor signaling and nuclear factor-kappaB, hypoxia-inducible factor-1A, and ETS transcription factors were all significantly enriched in ruptured IA[151]. However, there were also other studies that have made different observations[189; 194], for instance, genes coding for MMPs, genes coding for adhesion proteins of the extracellular matrix (ICAM1) and cytoskeleton (WIPF1, TUBA4A) were reported to be significantly under-expressed in the blood of ruptured IA patients[195], suggesting the mechanism of rupture is not regulated by degenerative process of the arterial wall.

MicroRNAs were also found to be potential biomarkers of IA formation and rupture. Differential expression of microRNAs have been linked to apoptotic cells and vascular wall activation pathways[196], phagocytes migration, mononuclear leukocytes proliferation[197], vascular smooth muscle cell modulation and even protein translation machinery[198].

A recent RNA sequencing (RNA-seq) study identified 229 differentially expressed genes in IA tissues by comparison to controls, and 1,489 genes were also found to be expressed in ruptured IA tissues by comparison to unruptured wall tissues[199]. The pathways of the differentially expressed genes in IA includes extracellular matrix (ECM), transmembrane transporter activity and blood vessel regulation, with top overexpressed genes as collagen *COL10A1*, cartilage intermediate layer protein *CILP2*, secreted frizzled-related protein *SFRP2*, RNA-binding family

member *MEX3B*; and underexpressed genes as *FAM134B*, transporter protein *SLC13A3*, coagulation gene *SERPIND1*, growth regulation gene *GREB1* and gap junction protein *GJB6*.

Several meta-analyses on IA gene expression have summarized the genes that were observed to be involved in independent studies. It was suggested that *BCL2, COL1A2, COL3A1, COL5A1, CXCL12, TIMP4* and *TNC* showed differential expression in more than three independent studies[200]. Another study using weighted gene co-expression network analysis suggested that *FOS, CCL2, COL4A2* and *CXCL5* enriched in immune response and extracellular matrix pathway contribute to the rupture of IA[201].

Recent studies also focused on expression levels on previous IA candidate genes discovered in previous GWAS and within functional CNVs. A CNV containing oxidoreductase *WWOX,* which previously was associated with cancer development, was found have significantly lower expression in IA tissues in Chinese[202]. Two GWAS significant genes *CDKN2A*[192] and *SOX17*[203] were both found to be downregulated in IA tissues.

It is interesting to note that there were large overlaps in the genes showing a differential expression profile in blood samples of IA[204], ischemic stroke[205] and AVM[206], especially with genes in the pathways of immune function and hematopoiesis, suggesting a potential continuity in the genetic etiology of different cerebrovascular diseases.

Genome-wide association study in IA risk gene identification

Following the extensive research of microarray expression data, genome wide association studies (GWAS) came into focus for the next stage of IA genetic research, using a dense SNP array (500,000 to 1,000,000 SNPs) covering the whole genome; an approach which is still deemed to be

the most successful method for the identification of IA risk gene. The first GWAS of IA was conducted in 2008[207], which comprised 2,100 IA patients and 8,000 controls across a discovery cohort of Finnish and Dutch, and a Japanese replication cohort. The study discovered significant signals on 2q (*PLCL1*), 8q (*SOX17*) and 9p (*CDKN2A*). A follow-up study increased number of European and Japanese IA study cohorts (5,891 cases and 14,181 controls) and identified three new loci: 18q11.2 (*RBBP8*), 13q13.1 (*STARD13*) and 10q24.32. The previous GWAS loci of 8q11.23-12.1 (*SOX17*) and 9p21.3 (*CDKN2BAS*) were also replicated, and the potential risk genes in those loci are involved in cell-cycle progression[208]. It is noteworthy that the 9p21.3 (*CDKN2BAS*) gene cluster has been identified as a GWAS hotspot for many other common diseases (e.g. coronary diseases, T2D, gliomas and basal cell carcinomas)[209].

After these two major GWAS, several small studies mainly focused on Japanese cases were conducted. A GWAS on Japanese IA cohort discovered other risk loci that related to the ELN/LIMK pathway, namely 9q31.3 (rs1930095, $p=1.31\times10^{-5}$), *TMEM195* (rs4628172, $p=1.32\times10^{-5}$), *ARHGEF11* (rs7550260, $p=4.93\times10^{-5}$) and *IQSEC1* (rs9864101, $p=3.63\times10^{-5}$)[208]. The additional risk loci of the second GWAS were also followed up in two independent Japanese IA cohorts and a new loci 4q31.23 (rs6841581, $p=2.2\times10^{-8}$) near the *endothelin receptor type A* (*EDNRA*) was highlighted in the discovery, along with the increased evidence of two additional loci 12q22 (rs6538595, $p=1.1\times10^{-7}$) and 20p12.1 (rs1132274, $p=6.9\times10^{-7}$)[210]. Association with genes *EDNRA* (4q31.22) and *CDKN2BAS* (9p21.3) were replicated in another larger Japanese GWAS (rs6842241, $p=9.58\times10^{-9}$; rs10757272, $p=1.55\times10^{-7}$), and the functional assessment of *EDNRA* SNP rs6841581 showed that it affects the expression of *EDNRA*[211]. A study following previous IA GWAS risk loci and found that among which *PRDM6* (5q23.2) was significantly associated with elevated blood pressure in Europeans ($p=8.13\times10^{-7}$)[212].

A meta-analysis study on all major GWASs further concluded that the IA consistent risk loci are 8q11 (rs10958409, $p=1.78\times10^{-15}$), 9p21 (rs10757278, $p=1.59\times10^{-13}$), and 4q31.23 (rs6841581, $p=1.95\times10^{-8}$), followed by 12q22 (rs6538595, $p=1.12\times10^{-7}$), 20p12 (rs1132274, $p=8.29\times10^{-7}$), 2q33 (rs1429412, $p=1.07\times10^{-6}$) and 7q13 (rs4628172, $p=4.04\times10^{-3}$)[179]. The GWAS risk SNPs rs6841581 (4q31.23), rs10958409 (8q11.23), rs9298506 (8q12.1), rs1333040 (9p21.3), rs12413409 (10q24.32), rs9315204 (13q13.1), and rs11661542 (18q11.2) were also investigated in another study of genetic risk scores (GRS) based on the site of IA; the results showed that GRS of IA were higher at the MCA in both Dutch and Finnish and combined cohort. However, the GRS of previous GWAS loci were not associated with familial IA in all cohorts, indicating that risk loci of familial IA may be different from the results of previous GWAS results[213].

Correlated with the GRS finding, another GWAS study focusing on populations of European descents was conducted with the inclusion of familial and sporadic IA cohorts. The two discovery cohorts yielded different risk loci, but meta-analysis confirmed the previous risk loci in 9p21.3 (rs6475606, $p=3.59\times10^{-8}$) and 8q (rs1072737, $p=8.7\times10^{-5}$). The suggestive loci of familial discovery cohort were different to any previously reported loci: *PDE1A* (rs1897472, $p=6.66\times10^{-7}$) and *BTBD16* (rs911774, $p=6.69\times10^{-6}$); however, the two loci were not significant after the inclusion of sporadic discovery cohort, which suggests again that familial IA may have different underlying genetic causes[214].

Focusing on population specific, low frequency variants, an IA GWAS on Finnish familial IA discovered high IA risk loci with low-frequency (MAF<0.05) variants on 2q23.3 ($p=1.42\times10^{-9}$), 5q31.3 ($p=3.17\times10^{-8}$), 6q24.2 ($p=1.87\times10^{-11}$) and 2q33.1 ($p=1.87\times10^{-12}$), while 7p22.1 was associated with the number of IAs[89]. Other GWAS focused on different populations also yielded new risk loci. A recent one focused only on European descendants identified new IA risk loci on

7p21.1 (rs10230207; $p=4.14\times10^{-8}$)[215], close to the SNPs previously associated with ischemic stroke[216]. However, this region failed to show replication in the Finnish cohort, suggesting the genetic heterogeneity of IA may be linked with population diversity. Subsequently, another multi-staged GWAS in Portuguese IA also suggested new loci in their population - rs4667622 (2q31.1, $p=7.13\times10^{-5}$), rs6599001 (3p22.2, $p=6.05\times10^{-5}$), rs3932338 (5p14.2, $p=2.02\times10^{-4}$), and rs10943471 (6q14, $p=5.50\times10^{-4}$)[217], further lengthening the list of GWAS risk loci, however this study had limited power due to its sample size.

Despite of the successful findings of IA GWAS, the accounted heritability in those GWAS risk SNPs is extremely low. For instance, in the Finnish study[89], all five newly discovered risk loci accounted for only 2.1% of the heritability. Additionally, the six SNPs identified in the previous GWAS study explained only 2.5% of the heritability in the Finnish. Besides the possibility of lack of power, all these results may also suggest that the genetic etiology of a heterogeneous disease group such as IA could be population specific and/or differs in familial form.

IA risk gene discovery using whole exome sequencing (WES)

The implementation of high-throughput sequencing and WES, which started in 2010, opened a major era of genetic research and has led to many successful discoveries, especially in rare Mendelian disorders. As a systematic way of exploring rare coding variations, the success rate of WES relying on how much of the disease can be explained by those variants and how many of the diseases exhibit Mendelian inheritance. Despite its wide application in disease gene discovery, the number of WES studies in IA is surprisingly limited, with only two such studies were reported up to date.

The first WES study of IA was reported in 2015 which examined Japanese families[218]. Twelve families with a total of 42 cases were included in the discovery cohort and validated in more familial and sporadic IA cases. The study identified 78 risk variants; and 10 variants from 9 genes (*GPR63, ADAMST15, MLL2, IL10RA, PAFAH2, THBD, IL11RA, FILIP1L,* and *ZNF222*) were further prioritized based on their functions. These ten variants were genotyped in a validation cohort and only the variant in *ADAMTS15* (rs185269810, p.E133Q) showed a significant association with familial IA (p=0.001). Functional assessment of this variant also highlighted its role in endothelial migration. A subsequent and independent WES study that focused on seven IA families of European American descent identified 68 potential risk variants[219]. Linkage analysis and RNA sequencing were applied as additional steps in order to further filter variants. Disease segregating variants were observed in 8 genes (*KLF11, ABCC3, TANC2, ALMS1, ARHGEF17, SMEK2*, *HTRA2*, and *NDST1*) and RNA-seq only revealed *TMEM132B* to be differentially expressed. Furthermore, none of the candidate variants were located in genes reported in the previous GWASs.

It is not surprising that there is no replication between the two IA WES studies. Genetic heterogeneity, phenocopies, or gene-environment interactions are all likely be factors that introduce complexities to WES discovery efforts. IA being such a common disease with many underlying factors that contribute to its predisposition, based on the findings from large amounts of data that generated by previous genetic studies (especially GWAS and WES), it is naturally to assume that a successful design of IA genetic study needs to rely on increasing the effect sizes of potential risk variants. As a result, the selection of a particular founder population with high IA prevalence and familial aggregation might lead to the enrichment of such variants with a higher heritability.

1.4: RATIONALE AND OBJECTIVES

Since the very beginning of applying genome-wide sequencing and genotyping methods in the complex disease genetic studies, the collection of large cohorts has been critical. However, an alternative approach involving founder populations also works well to increase the power of detection of disease genes. IA is such a complex disease with a low penetrance that it can be affected by environmental factors and genetic background. Moreover, in the case of familial aggregations, certain populations seem to present higher genetic predispositions[110; 211].

Many two-staged studies have successfully combined the uncovering genetic architecture of founder populations with risk gene discovery to the disease predisposed in such populations; this yielded significant findings in the two areas: such as in Ashkenazi Jewish peoples with Crohn's disease[220; 221], Greenlanders with T2D[14; 30], Sardinians with multiple sclerosis[222; 223] and Indians with cardiomyopathy[224; 225]. French-Canadians and Nunavik Inuit are two important, yet very distinctive, founder populations in Quebec, a high prevalence of SAH and IA is observed in both. The genetic causes behind this increased IA prevalence are not well understood, while in the case of Nunavik Inuit, their underlying genomic landscapes are equally intriguing. In order to shed some light on these questions we conducted a series of studies which are described herein.

The first objective of this thesis was to reconstruct the genetic history and uncovering genomic signature of adaptations in Nunavik Inuit. For this purpose, we have used a WES and Illumnia Omni SNP-chip tandem approach on a cohort of Nunavik Inuit and performed analyses regarding their genetic structure, admixture, patterns of migration and natural selection. The genes located in the selection signals were further assessed by qPCR and tested for associations between

Inuit individuals affected with IA and controls. This study aimed to reveal the genetic architecture of a Canadian aboriginal population while also attempted to apply results from the population genetics analyses to a complex disease association. Subsequently, the genetic burdens of a family of lipid metabolism genes were also investigated in Nunavik Inuit, including *CPT1A*, which harbor the most significant variant under selection sweep in the Arctic indigenous population. The combination of the two methods (WES and SNP-chip) has helped to provide a full picture of the genomic landscape of Nunavik Inuit, which can be used in future studies aiming at analyzing other diseases for which this population is observed to be predisposed.

The second objective of this thesis was to identify genetic factors that could explain some of the missing heritability of IA in the French-Canadian founder population. The first part of this study was mainly high-throughput sequencing (HTS) based. WES data was generated from a discovery cohort comprised of large French-Canadian families in which multiple members affected with IA. Subsequently, genes that presented an increased burden of rare, functional variants in IA patients were re-sequenced in a larger French-Canadian replication cohort, with newly discovered potential risk variants from the candidate gene, and further assessed using functional assays. Our hope was that using a strategy that both minimize the population heterogeneity and acquiring unbiased genetic data would increase the power of identifying the genetic risk of IA. Moreover, we complemented the French-Canadian WES study by using high density SNP-chip to perform a GWAS analysis for familial IA cases and a large cohort of French-Canadian control individuals. This complementary approach have enabled us to identify low frequency variants associated with IA that would have been overlooked by previous GWAS focused on mixed populations.

CHAPTER 2: THE GENETIC SIGNATURES OF NUNAVIK INUIT

2.1: PREFACE

Nunavik Inuit represent an important and unique population. The harsh Arctic environment has shaped their genomes to harbor unique genetic variants that have enabled their adaptation. They are also the descendants of the third wave of human migration into the New World and, unlike other Native Americans, they have remained largely isolated until very recently. However, the recent changes towards a western lifestyle are thought to have brought increased prevalence of cardio-/cerebrovascular diseases among them, including IA, and genetic risks factors lay undiscovered in their unique genomes. Until recently, population epidemiological and genetic studies aiming to identify genetic risk factors were limited to studies examining disease candidate genes or focused on mtDNA and Y-chromosomal DNA. With the recent advances of high-throughput sequencing (HTS) technologies and the completion of 1KGP phase III data, which has included high quality whole genome sequencing data of 26 present-day populations, genome-wide population genetic studies have become feasible.

Here we have used a parallel HTS and genome-wide SNP-chip genotyping approach to obtain unbiased genetic data from Nunavik Inuit. Chapter 2.2 presents the first study characterizing the genetic history and structure of the Canadian Inuit population. Interestingly this study yielded findings different from those of an earlier one, where a neighbouring Greenlandic Inuit population was examined[14]. Unlike their Greenlandic counterpart, Nunavik Inuit have almost no European admixture, and some even harbor ancestries that closely resemble those of the ancient Native

Americans. These results not only indicate that Nunavik Inuit is an old and isolated population, but also suggest that a higher percentage of their genome is unique to the major worldwide populations. However, while the Nunavik Inuit genome became adapted to a harsh environment and living conditions, it also brought about some genetic predisposition to certain "new" diseases. The evidence for this came from both the identification of selection signals, which are essentially embedded within the coding portion of the Nunavik Inuit genome and expression level analysis of the genes harboring variations of these selection signals[29]. These genes, which encode protein products associated with various biological processes (e.g. cell adhesion and immuno-response), and deemed to be indicators of the predisposition to cardio-/cerebrovascular conditions. Moreover, an association study aiming to identify variants under selection regions and IA identified a common variation in *CCM2*, which showed a significant association when in a homozygous state.

Chapter 2.3 focuses mainly on estimations of accumulated genetic burdens to profile *CPT1A* as well as other members of this gene family in Nunavik Inuit. The *CPT1A* p.P479L loss of function variant was initially discovered in Inuit who had CPT-I deficiency[52] that was linked to infant mortality[226]; it was subsequently determined that it is within a selective sweep particularly in Arctic indigenous peoples[40]. We found that the frequency of this variant has increased with the proportion of Inuit ancestry, and in Nunavik Inuit, the deemed derived allele of p.P479L reached a near fixation of 96%. We also found increased mutation burden in other genes encoding carnitine acyltransferases including *CPT2* and *CRAT*, in Nunavik Inuit compared to the Asian and European populations, highlighting the importance of fatty acid oxidation in Nunavik Inuit's adaptations.

Overall this chapter presents a series of studies of reconstructing genetic architecture of Nunavik Inuit, in which we have addressed multiple points, from the population genetic struture and admixture to migration to genes positively selected for the adaptations. The studies have also

suggested the possibilities of identifing disease risk genes based on population genetic approaches. Such methods might be applied in different populations as alternative means to carry out genetic study of complex diseases.

2.2: THE GENETIC HISTORY AND SIGNATURES OF NUNAVIK INUIT

Manuscript: The genetic history and signatures of Nunavik Inuit

Authors and affiliations:

Sirui Zhou[1,2], Pingxing Xie[1], Amelie Quoibion[1,3], Amirthagowri Ambalavanan[1,3], Alexandre Dionne Laporte[1,5], Dan Spiegelman[1,5], Cynthia V Bourassa[1,5], Lan Xiong[2,4,5], Simon Gravel[3], Patrick A. Dion[1,5], Guy A. Rouleau[1,5]

1) Montreal Neurological Institute and Hospital, McGill University, Montréal (Que), Canada; 2) Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal (Que), Canada; 3) Department of Human Genetics, McGill University, Montréal (Que) Canada; 4) Centre de recherche, Institut universitaire en santé mentale de Montréal (Que) Canada; 5) Department of Neurology and Neurosurgery, McGill University, Montréal (Que) Canada.

Authors contribution:

Study design: SZ, LX, SG, PAD, GAR

Experiment: SZ, AQ

Bioinformatic analysis: SZ, AA, PX, ADL, DS

Sample recruitment: CVB

Manuscript writing: SZ, LX, PAD, GAR

Abstract

The Canadian aboriginal population has a distinct history and some health risks are known to specifically affect its individuals. Nonetheless the number of genetic studies examining this population is limited and much remains to be uncovered regarding its genetic characteristics. In this study we generated whole exome sequences and genome-wide SNP-chip genotypes for 170 Nunavik Inuit that belonged to a small and isolated, yet important, founder population of Canadian Arctic indigenous people. Our results revealed that the genetic background of Nunavik Inuit is distinct from the ones of present-day world-wide population. The majority of Nunavik Inuit are with un-admixed Inuit ancestry; only individuals from the Ungava Bay villages have certain admixtures found with the ancient Native Americans. The Inuit live in the Hudson Bay area are also seemly distinct than the ones located next to Ungava Bay; and the colonization of Nunavik appear to have consisted of two southward migrations from the Northwest corner, neighboring Nunavut. Furthermore, we observed certain genomic signatures that are distributed across different loci of coding regions which are supportive of the adaptation pressure unique to Inuit. The strongest selection signal and the significant level of gene expression changes are related to genes involved in cell adhesion and immune response processes (*CPNE7, ICAM5, STAT2* and *RAF1*), which suggests adjustment of their functions during the development of the cardiovascular system. A subsequent analysis aimed to establish if any association could be observed between these signatures and the risk of intracranial aneurysms (IA) in Nunavik Inuit revealed a splice variant (rs2289367) in cerebral cavernous malformations gene (*CCM2*); with much higher frequency in Nunavik Inuit than general populations (0.83 vs 0.22), which was significantly associated with IA, especially in homozygous state (OR=0.23).

Introduction

Considering that the migration of Aboriginal Canadians occurring during the settlement of the Arctic population took place over a period of approximately 6,000 years[1], it is noteworthy that present-day Nunavik Inuit are the descendants of a long line of different occupants with genetic discontinuity among most of them. The occupation of this territory started with the early Paleo-Eskimos (at some point between 3,000 to 800 BCE), who subsequently replaced by Dorset culture (1,100 CE) when the Thule people began to occupy the territory[9] and became the ancestors of all modern Inuit. The different Arctic population changeovers have mostly followed distinctive climate shifts in history, and it is assumed that both environment and isolation have played a large part in shaping the Inuit genome. In the past decades, changes of the Nunavik Inuit's lifestyles were believed to have increased their risk to several complex disorders[227], such as cardiovascular diseases (CVDs)[84; 228], metabolic disorders[229], infection[71] and cancer[74].

A number of recent studies have genotyped the Siberians and the Greenlandic Inuit using high-density SNP arrays and genomic profiles of these arctic indigenous peoples have emerged. An analysis on cold adaptations of the Siberian populations revealed that the genes involved in lipid metabolism and vascular smooth muscle contraction were being selected[39], with strong selective sweep of a *CPT1A* deleterious variant described in a follow-up study[40]. Another study of the Greenlandic population subsequently revealed that there has been a substantial recent European gene flow into the Greenlandic Inuit, and that the populating of Greenland came from a single migration from northwest towards southeast[14]; Subsequently a strong selection signal was reported in a cluster of fatty acid desaturases (*FADS1-3*), which determine polyunsaturated fatty acid

(PUFA) levels[29]. Another team independently reported a *TBC1D4* founder mutation related to the increase of prevalence of T2D in the Greenlandic Inuit[30].

Despite these recent advances there are various aspects of the genetic profile of the Inuit populations that remain to be explored. While significant genetic divergences exist between the Nunavik and Greenlandic Inuit, the latter are nonetheless genetically closest to the Nunavik Inuit, in particular to the contribution of European gene flow. The underlying genetic factors potentially associated with higher prevalence of cardiovascular and related disorders in Arctic indigenous peoples are poorly understood[5; 230], especially in the evolutionary and genome-wide scale. Through a combined analysis of high density SNP-chip genotyping and whole exome sequencing (WES) across an Inuit cohort that sampled over 1% of the population of Nunavik, our study provides clues that will help to clarify questions about the susceptibility of the underlying genetic factors.

Results

Population structure of Nunavik Inuit

Principle Component Analysis (PCA) of the panel ALLPOP containing a total of 5,422 present-day and ancient individuals showed that Nunavik Inuit were largely separated from the other worldwide populations, while being the closest to the Siberian Eskimos and the ancient New World/New World Arctic populations (**Figure S1**). After exclusion of the Africans and other populations, which have little relation with Nunavik Inuit (panel NUIrelate), and with most of the ancient humans removed as outliers because of their low sequencing qualities, the PCA results suggest that the closest relatives to the Nunavik Inuit are the Eskimos and the Greenlandic Inuit. However, the Paleo-Eskimo (Saqqaq)[16] individuals are shown to be the closest to the Siberian

Koryaks and do not resemble much the Inuit from either Nunavik or Greenland (**Figure 1A**). 146,668 SNPs within the high-confidence (HC) regions were derived from the WES data of the Nunavik Inuit and the publicly available WGS data of 649 individuals, including four Native American populations from the 1000 genome project (1KGP) and several indigenous populations and ancient individuals described in **Table S1**. CHB and CEU from the 1KGP were also included to help determining the ancestries. With further elimination of possible SNP with ascertain bias, the Nunavik Inuit still appear to have a distinctive ancestry when compared to other populations, including the Greenlandic Inuit and the Northeastern Siberians (**Figure 1B**). There were also slight genetic differentiations within the Nunavik: the two Ungava Bay villages (Kuujjuaq and Kangiqsualujjuaq) seem to be separated from the rest of the villages located near Hudson Bay (**Figure 1C**).

ADMIXTURE analyses[231], using the same WES HC dataset, similar results were obtained. When assuming the scenario of four estimated ancestries (Han, Native-American, European and Inuit), individuals from the Ungava Bay villages seem to contain Native American ancestry admixture: among 27 Inuit individuals from Kuujjuaq and Kangiqsualujjuaq, six of them were found to have more than 15% of the Native American ancestry (**Figure 2A**). We preformed further analyses using the SNP-chip dataset (NUIrelate) which contains a total of 103 world-wide populations. When the number of estimated ancestries was increased to 8 (Euro1, Euro2, Central Asian, Siberian, Han, ancient New World, Native-American, Inuit), we found a major non-Inuit ancestry that existed only in the Ungava Bay villages, with an average of 16% of admixture in the Kangiqsualujjuaq Inuit. Interestingly, this admixed ancestry was found mainly in the genome of ancient humans, mostly from the New World Arctic (**Figure 2B)**. Furthermore, 21 out of 170 Nunavik Inuit showed European admixture (>1%), which seems to be introduced by very recent

admix events (10 individuals have >25% European ancestry). Conversely, 87% of the Nunavik Inuit showed no evidence of European admixture. Our results also showed that the present-day Inuit ancestry first started appearing around Altai regions and then increased drastically from the Western Siberia to Northeastern Siberia, which was significantly decreased in the Native American populations. Only several groups of Aboriginal Canadians (Athabascans, Algonquins, Chipewyans, Crees and Ojibwas) still have some admixture similar with the Inuit ancestry (**Figure 2B**).

PCAdmix[232] was used to infer genome-wide local ancestries of admixed Inuit individuals and Arctic indigenous peoples. The Inuit individuals from Kangiqsualujjuaq were estimated to have an average of 2.0-3.8% of their ancestry from the ancient New World populations, inferred from genomes of Saqqaq, Clovis[233] and Kennewick[234] (**Figure 3**). The Saqqaq individual shared 67% of the ancestry with today's Nunavik Inuit; however, the ancient ancestry inferred from the ancient Arctic genome[9] only accounted for 6.9% of the Kangiqsualujjuaq genome, suggesting the low sequencing quality of the ancient Arctic genome may have caused bias in the ADMIXTURE inferring process (**Figure S2**). Among the 21 Nunavik Inuit who have European admixture, the European ancestry accounted from 12.2% to 97.6% of their genomes. In other aboriginal populations, Greenland Inuit shared the most of the ancestries with Nunavik Inuit (83.4%), followed by Naukan (72.7%), Eskimo (64.4%), Chikuchi (46.8%), Chipewyan (26.3%), Koryak (24.3%), Algonquin (12.3%), Evens (11.7%), Yukagir (9.75%), Cree (9.5%) and Ojibwa (9.1%) (**Figure S2**).

History of populating of Nunavik

To investigate the effect of isolations on the Nunavik Inuit, we estimated the linkage disequilibrium (LD) in individuals from several Nunavik Inuit villages and compared them with other Arctic populations (Naukan, Eskimo, Chukchi, Western and Eastern Greenlanders) and aboriginal Americans (Chipewyan and Athabaskan). Because sample size affects the accuracy of LD estimation, we only included samples from five villages (Inukjuak, Puvirnituq, Ivujivik, Salluit and Kaniqsualujjuaq) which have more than 15 individuals. The results showed Nunavik Inuit had the highest degree of LD compared to the Arctic indigenous peoples and the 1KGP major populations, and they also had significantly higher LD than the Greenland Inuit. Since the presence of gene flow reduces LD, we concluded that Nunavik Inuit is the most historically isolated population in the Arctic indigenous peoples, including Greenlandic Inuit. Within Nunavik, the Kangiqsualujjuaq villagers had a higher LD compared to the Inuit from the Hudson Bay villages, suggesting it was the most isolated sub-population in Nunavik Inuit (**Figure 4**). These observations are correlated with the result of the runs of homozygosity (ROH) analyses, in which 27.7% of the ROHs are longer than 10 Mb in the Kangiqsualujjuaq Inuit (**Figure 5**). The Inuit from Salluit and Kuujjuarapik have a much smaller percentage in the long segments of ROH, which may be due to the inclusion of several highly European-admixed individuals.

In the Treemix[235] analysis, Yoruba was set as outgroup in rooting the tree to infer the maximum-likelihood genetic-drift topology relating the Nunavik villages to other artic populations. The result of the tree shows the root split between Nunavik Inuit and Greenlandic Inuit, with the two Inuit populations subsequently split off with the New World Arctic indigenous peoples and the Northeast Siberians. There was also a split between the Ungava Bay Inuit and the Hudson Bay Inuit. A potential migration route from the Salluit to both the Southwestern villages and the Ungava Bay villages along the coastline could also have be established, with the Inuit from

Salluit seemed to be the closest to the New World Arctic indigenous peoples, suggesting the ancestors of Inuit entered Nunavik at the northwest point from Nunavut (**Figure 6**). Furthermore, there were migration events from Kangiqsualujjuaq to Kuujjuaq, indicating the history of connections between these two villages. No evidence of migration events appeared in other Nunavik Villages, while many occurred in aboriginal Canadians, correlated with our previous findings of little admixture existed in Nunavik Inuit.

Genes under natural selection in Nunavik Inuit

Population branch statistics (PBS)[236] were calculated for a total of 387,339 single nucleotide variants (SNVs) generated from WES data between the Nunavik Inuit and the CHB with the CEU as outgroup. 38,116 SNVs with *PBS* >0.1 was shown in a Manhattan plot (**Figure 7)**, from which 9,883 were coding SNVs (red), 784 located in or near splicing sites, 1,936 were in UTRs and the rest were non-coding, intronic or intergenic SNVs. Variants with an empirical threshold of 99.9th (*PBS*>1.0) were selected as potentially under natural selection. 132 SNVs from 79 genes met the aforementioned criteria, including 38 coding and UTR SNVs (**Table S2**). In result, rs80356779 in *CPT1A* (p.P479L) showed the highest PBS score (3.11), which correlated with our previous study[49]. This Arctic-specific allele had a frequency of 68% in the Northeastern Siberian populations[40]. It was also absent in the ancient Siberian individual Mal'ta[56] from 24-kya BP and two ancient Native Americans: the Clovis from 11-kya BP and the Kennewick man from 8-kya BP. However, the SNV was presented as heterozygous in current day Nivkhs, Athabascans and Aleutians[237], as well as the 4-kya Saqqaq and a late Dorset individual 1.6-1.4-kya BP. It could therefore be assumed that this SNV occurred in the Neolithic Siberia after the first wave of

migration into the New World, and it only reached an even higher frequency around the Arctic sea.

Besides *CPT1A*, we also confirmed the selection signals previously found in the Greenlandic Inuit[29] around the *FADS* cluster (chr11: 61,543,499- 61,647,427), which included *MYRF*, *TMEM258*, *FADS1*, *FADS2* and *FADS3*, with the highest PBS at 1.17 in *FADS1*. Several new regions with at least one SNV highly selected (*PBS*>1.5) and within the 500 kb window included chr16: 89.6 Mb- 90.1 Mb (*CPNE7, DPEP1, CDK10, TCF25 and PRDM7*, with the highest PBS at 1.74 in *CPNE7*); chr3: 12.6 Mb- 12.9 Mb (*TSEN2*, *RAF1*, *MKRN2* and *CAND2*, with highest PBS at 1.76 in *TSEN2*); chr8: 143.9 Mb- 144.0 Mb (*GML, CYP11B1* and *CYP11B2* with highest PBS at 1.52 in *CYP11B1*) and *IMPDH1* at chr7 with highest PBS at 1.65. A single missense SNV in *STAT2* had also been highly selected (*PBS*>1.53), however it was the only SNV found in this region (**Figure 7**).

According to the previous selection signals based on the WES data of the Greenland Inuit, there was a strong correlation between the Nunavik and the Greenland Inuit populations. Among 132 SNVs (79 genes) with *PBS*>1.0 that we found in Nunavik Inuit, 61 were also found in Greenland Inuit with *PBS*>0.3, from which 25 were also being strongly selected (*PBS*>1.0).

We also calculated the WES-based PBS between the Nunavik Inuit and the Northeastern Siberians with CHB as outgroup to detect more recent selection. One missense SNV (rs180768267) had the highest PBS located in *CAND2* (*PBS*=1.09), followed by *ATP10D* (rs16851681, *PBS*=1.07) and *ICAM5* (rs1056538, *PBS*=1.04). Both SNVs in *CAND2* and *ICAM5* were predicted to be deleterious according to their combined annotation dependent depletions scores (CADD >15), suggesting the SNVs may have important function in certain biological progresses. However, the *CPT1A* rs80356779 SNV was only moderately selected between the

Nunavik Inuit and the Northeastern Siberians (PBS= 0.84), followed by the aforementioned missense SNV in *STAT2* (PBS=0.84); both of them were predicted to be highly deleterious (*CPT1A*, CADD=18; *STAT2*, CADD=25) (**Table S3**).

RT-qPCR was used to test the expression levels of genes under selection between the Inuit and controls who live in close geographic area but without selection, i.e., the French-Canadians (FCs). Thirty genes (**Table I**) were selected from the SNVs with top PBS and met the following criteria: 1) 2 or more SNVs have *PBS*>1 and in the same region or located in the abovementioned selected regions; 2) the gene also showed sign of selection (*PBS*>0.5) in Greenlandic Inuit[29]; and 3) the gene expression in lymphoblastoid cell lines (LCL) cells is predicted to be measurable by GTEx (www.gtexportal.org). Two genes, *EDAR* and *SLC24A5* were previously shown under selection among worldwide populations, which therefore were not tested in this study[238].

Among the genes tested, *GML, CYP11B1, CYP11B2* on chr8, *IMPDH1*, *NCR1*, *DSP, LECT1* and *IGHMBP2* showed very limited expression in LCL, therefore were excluded from the subsequent analyses because of their low confidence in the expression measurements. Within the rest of 21 genes, 11 of them showed the same trend of expression in relative quantification (RQ) compared with the FC controls across three independent experiments. Two genes in particular, *CPNE7* (P = 0.045, Wilcox test) and *ICAM5* (P = 0.041, Wilcox test), had significant differences in the expression levels (measured by delta CT) in the Nunavik Inuit compared with the FCs; while the level of expression of another three genes *STAT2* (P = 0.051, Wilcox test; P = 0.03, T-test), *RAF1* (P = 0.063, Wilcox test; P = 0.05, T-test), and *FADS1* (P = 0.051, Wilcox test) suggested borderline significance for each gene (**Figure 8**).

GO enrichment test on 1,596 genes with one or more coding SNVs with weak signals of selection (PBS>0.3) revealed significant enrichment of biological processes in cell adhesion (GO:0007155) (P = 6.52e-03); localization (GO:0051179) (P = 3.79e-03) and response to stimulus (GO:0050896) (P = 7.46e-05) (**Table S4**), from which 124 genes are categorized under cell adhesion process, including the *ICAM* family genes*, ICAM-1 to ICAM-5,* which showed the top selected signals.

Forty-nine SNVs from the WES data predicted to be both protein-damaging (missense and splicing) and also under potential selection in Nunavik Inuit (*PBS*>0.8). The result of the association tests of these 49 SNVs between the Inuit IA patients and the matched population controls revealed three SNVs spanning 35kb on chr7 to be significant, with one SNV in the gene *CCM2* remained to be significant after multiple correction (rs2289367, p=0.000952) (**Table S5**).

Discussion

In this study, we combined the datasets of WES and dense SNP-chip genotyping and applied statistical methods developed for population genetics; and then investigated the genetic history and signatures of Nunavik Inuit. To our knowledge the WES dataset presented here is the largest one generated for a single indigenous population. Moreover, the combination of this WES data to high quality genome-wide SNPs genotyping data represents an unbiased approach to study the largely unknown genetic profile of this unique population. We also associated the Inuit genetic profile with a population-characteristic disease phenotype as a portion of the Nunavik Inuit samples examined here were ascertained as IA patients. IA is a complex cerebrovascular disorder, where there is a thinning and bulging of the intracranial vascular walls, and the prevalence of IA is increased in the Nunavik Inuit and the Canadian aboriginal peoples. Our results suggest that the

distinctive genetic signature of the Nunavik Inuit population is the result of natural selection or genetic drift, and some element of this signature could also explain the increased IA risk.

Genetic structure and migration history of Nunavik Inuit

Our observations show that the isolated Nunavik Inuit population has no admixture from other ancestries and is distinctive from other Arctic populations, such as Greenlandic Inuit, Alaskan Athabaskans, Canadian Chipewyan and Northeast Siberians. The component of Inuit ancestry is almost exclusive in Nunavik Inuit, which makes it an ideal "reference" population for future genetic studies of Arctic indigenous peoples. Apart from the fact that Nunavik Inuit are genetically different from other indigenous populations, there was also a genetic isolation between the Inuit living at the West Coast of Nunavik (Hudson Bay) and the East Coast (Ungava Bay), where Ungava Bay village Inuit seem to have part of their ancestry came from the 4-11 kya BP ancient New World. This evidence may suggest two scenarios: 1) Nunavik Inuit at Ungava Bay had once admixed with late pre-contact Innu of Labrador, whose ancestors were Paleo-Indians[9]; and 2) there had been genetic continuity between Dorset and Thule Inuit in Nunavik, which correlates with the findings of Helgason *et al.*[15], although the study in Greenland Inuit[14] suggested otherwise. Regardless of the conflict findings, there were evidences indicate that Dorset people remained in part of the eastern Arctic Canada until much later (1,350 CE), when Thule people already had occupied Greenland at 1,100 CE[239]. However, the percentage of this ancient Arctic ancestry is low (<5%), suggesting an ancient admixture. On the other hand, only 12% of the Nunavik Inuit have European admixture (>10%), with ten of them (out of 170 studied individuals) having high percentage European ancestry (>50%), which is correlated with the documented history that only until 19th century Nunavik had European encounters.

We also found evidence suggesting that Thule Inuit came into Nunavik through the Sugluk Inlet at Salluit, followed by costal migrations towards the east and west of Nunavik. The Greenlandic Inuit split off with the rest of Nunavik Inuit, correlating with the Thule migration split off at the Canadian high Arctic in 1300 AD into Greenland and Eastern Canada (Nunavik and Labrador)[9]. We have also discovered a closer split between West Greenland and Ungava Bay Inuit, which could be related with a small group of Inuit from Baffin Island arrived in the Northwest of Greenland in 1864[240], who might bear close genetic resemblance with the Ungava Bay Inuit.

Similar to that of Greenlandic Inuit, Nunavik Inuit have elevated LD; and this trend seems to increase from the West to the East. There was also a higher percentage of large ROHs (>10 Mb) in the Ungava Bay villages, which indicates higher levels of isolation and less admixture. The increased LD and higher levels of genetic differentiation (indicated by F_{ST}) in Nunavik Inuit may suggest an increased probability that deleterious variants may reach high frequencies due to many generations of genetic drift or hitchhiking with the nearby variants of natural selection.

Genetic signatures of Nunavik Inuit highlighting genes involved in cardiovascular risk

Genes involved in lipid metabolism are known to be under positive selection in Arctic indigenous populations, especially in Inuit[29; 40]. In addition to confirming the two regions (*CPT1A*, *FADS1-3*) on chromosome 11 which were previously identified to have high level of selection signals, we have also discovered additional loci with unique genetic signatures in Nunavik Inuit, by mainly focusing on the protein coding regions of their genomes.

A strong selection signal in the Nunavik Inuit was identified within a 105 kb region on chr16 (chr16: 89,656,156-89,760,474), in which variants in three genes *CPNE7, DPEP1* and

CDK10 showed the highest PBS. *CPNE7* is the most prominent gene, which showed significant expression level changes ($p<0.05$) between LCL cells of Nunavik Inuit and FCs (European descents); and had another variant also being positively selected in Greenlandic Inuit (rs139901937, $PBS^{GI} = 1.39$). An alternative splicing variant rs12445560 (*PBS*=1.74) in *CPNE7* had an extremely high frequency in Nunavik Inuit (*F* = 95%), while gradually decreasing in Northeast Siberians (*F* = 74%) and Native Americans (*F* = 40%). The frequency of rs12445560 is much lower in Europeans and East Asians (*F* = 10-12%) and even rarer in Africans (*F* = 1%). The frequency changes of this splicing variant is interesting and seems to be correlated with the out-of-Africa migration, and has reached near fixation in Nunavik Inuit. CPNE7 is a member of the copine family that bind phospholipids in a calcium-dependent manner and also belong to von Willebrand factor A-domain-containing proteins[241], which is known to be important in cell adhesion and migration events. Copines are a family of protein with similar structure and biological functions, although their functions are not fully described, there were studies suggesting that variants in *CPNE4* may be associated with coronary artery disease in East Asian population[242].

Another locus with significant selection signal is located on chr19, containing genes *ICAM5* and *ICAM1* (intercellular adhesion molecules). In our experiment, *ICAM5,* with two missense variants (in LD) under selection, showed significant changes in gene expression between the Nunavik Inuit and FC controls. The common haplotype of these two SNPs also seems to be present in higher frequency in the indigenous populations (0.19 in CHB, 0.38 in CEU, 0.6 in Native Americans, 0.56 in Northeast Siberians), and has again reached near fixation in Nunavik Inuit (0.98). ICAM5 is neuronal specific, and ICAM1 is a critical molecule secreted by endothelium during vascular inflammation and is responsible for formation of atheroma[243]. *ICAMs* also encode a family of adhesion proteins, which are involved in oxidative stress and inflammatory response,

while highly expressed in the brain and being crucial in the endothelium remodeling, make them potentially very important players in cardio- and cerebrovascular diseases[244-246].

Other interesting genes with variants under selection in the Nunavik Inuit include *STAT2,* which encodes a critical component in the type 1 interferons signalling[247], and is a well-studied inflammatory mediator and important factor in antiviral defense, especially in the brain[248]. Other studies indicated STAT2 is a prominent modulator of inflammation in vascular cells and during atherosclerosis[249], which makes it a potential marker for CVD[250]. *RAF1* also encodes an inflammation regulator and activated by *fibroblast growth factor* (*FGFB*) and *vascular endothelial growth factor* (*VEGF*)[251], making it critical in endothelial cell survival during angiogenesis[252]. *RAF1* expression was seen to be altered in patients with arterial hypertension[253]. The selection signal on chr8, containing genes *CYP11B1-CYP11B2*, which function in steroidogenesis and are known to be involved in hypertension and increased cardiovascular risk[254].

Besides *FADS1*, four other genes, *CPNE7*, *ICAM5*, *STAT2* and *RAF1,* which contain variant(s) most likely under positive selection, and which also have shown high levels of differential expression in the Nunavik Inuit and the FC controls, all function in cell adhesion and inflammatory response, suggesting that they are also possibly involved in the cardiovascular development. The variants, which have reached near fixation (>80%) in Inuit, were predicted to be highly damaging according to their CADD scores, thus their impact on the proteins' functions and the metabolisms they act upon might be differently in Nunavik Inuit by comparison to other populations. Genes involved fatty acid metabolism were also known to be associated with the increased risk of CVD[255], which have shown the most positively selected signals in the Inuit populations. However, we have discovered genes associated with vascular remodelling and inflammatory response – other features of CVD could also be associated with the altered frequency

and the differential expressions in Nunavik Inuit. The accumulation at high frequency of potentially deleterious variants is usually due to less selective disadvantage in an isolated, small founder population with increased LD (genetic drift). However, because of the drastic frequency changes, the scenario that we discovered here resembles more of natural selection, in which these genetic signatures were essential in the adaptation to the Inuit living environment. Nevertheless, the recent switch to Western life style and diet may have diminished the beneficial effects that brought by those adaptations in the Nunavik Inuit and made them predisposed at high risk to cardiovascular diseases.

Splicing variant in CCM2 under strong selection is associated with the risk of IA in Nunavik Inuit

Intracranial aneurysm (IA) shared pathological similarities with some other cardiovascular diseases, such as the damaged vascular remodelling process and maladapted immunological responses. It was also demonstrated that lipid metabolic process could contributed to the pathogenesis of IA[256]. We hypothesized that among those highly selected genetic signatures, which we previously reported, some may also be associated with increased prevalence of IA in Nunavik Inuit.

As a result of an association test on variants which showed strong selection in Nunavik Inuit, a synonymous variant (rs2289367) in gene *CCM2,* predicted to affect canonical splicing, has shown to be significantly associated with the increased IA risk in this population. rs2289367 is common in populations worldwide, with similar frequencies from the lowest in East Asians (0.2) to the highest in Africans (0.26). However, the derived allele of rs2289367 is much more prevalent

in Inuit populations from Nunavik to Greenland (0.82 to 0.88). Furthermore, none of the Inuit IA patients were homozygous of the reference allele (G) and 86% of the IA patients were homozygous of the derived allele (A); while in controls only 53% were homozygous state (allelic OR=0.23). In comparison, only 1-4% of the general population are homozygous of the derived allele, which might possibly correlate with the overall worldwide prevalence of unruptured IA (3.2%)[96].

CCM2 is a well-established gene with mutations known to cause cerebral cavernous malformations (CCM), which are brain vascular malformations and characterized by abnormal enlarged capillary cavities without intervening parenchyma[257], likely to result in hemorrhage and consequent strokes. The two-hit mutation events leading to the complete loss of function in *CCM2*[258] are considered to be the main cause of CCM. In our study, most of the Nunavik Inuit IA patients were shown to have a homozygous, potential splicing variant on exon 8, which may lead to the complete functional changes of CCM2. CCM2 functions through the regulation of endothelial cell integrity[259]. A recent study suggested the gain of MEKK3 signalling, which CCM2 binds to, affects endothelial cell migration[260] and cardiovascular development[261]. The KO of *Mekk3* in mice is lethal due to multiple intracranial haemorrhages[262]. *CCM2*-knockdown human umbilical vein endothelial cells (HUVECs) has shown lack of the formation of lumens[263]; and the depletion of this gene was also known to induce a response similar to hemodynamic shear stress on endothelium remodeling[264] and vascular dilation with the thinning of vessel walls [265]. All of these observations seem to correlate with the well-accepted hypothetical pathogenic mechanisms and characteristics for IA[266], further suggesting the etiologic continuity between brain vascular disorders.

In summary, we have presented a systematic characterization of the genetic history, genomic signatures suggesting adaptations and the susceptibilities of IAs in Nunavik Inuit.

Utilizing of the WES data has fully taken account of the unbiased protein coding sequences of this important indigenous population. We have discovered new evidences that had not been revealed in previous Greenlandic Inuit studies. We have found that Nunavik Inuit is a unique ancestral population with potential minor Native American admixtures; and that they have distinct genetic signatures showing adaptations to their life styles and reflecting in their cardiovascular development in extreme environment; and that a splicing variant in the *CCM2* gene showing strong natural selection is also associated with the risk of IA. All of these have demonstrated the great value underlying the genetic and medical research in isolated founder populations.

Subjects and Methods

Subjects and genetic data

Nunavik Inuit

One hundred and seventy Inuit individuals from Nunavik (Québec, Canada) (**Figure S3**) were included in this study, 155 of them were coming from ten villages: 38 individuals from Kuujjuaq (9) and Kangiqsualujjuaq (29) were recruited from Eastern Nunavik villages around Ungava Bay and 117 individuals were recruited from Ivujivik (19), Salluit (23), Puvirnituq (28), Inukjuak (32), Akulivik (5), Kuujjuarapik (2), Umiujaq (3) and Whapmagoostui (5) in Western Nunavik villages around Hudson Bay. The rest 15 individuals were from unknown village origins. Written informed consent was obtained from all participants and the study was approved by McGill University Ethics Committee and the Nunavik Nutrition and Health Ethics Committee.

SNP-chip genotyping and WES for Nunavik Inuit

Illumina HumanOmniExpress 12 and 24 were used to genotype 165 Nunavik Inuit, each contain 730,525 and 716,503 SNPs, respectively; with WES performed on 114 selected genotyped Nunavik Inuit. Illumina GenomeStudio was used for the SNP-chip genotype calling. For exome sequencing data, steps regarding to the unaligned bam-to-fastq conversion by picard, adaptor trimming, alignment, sam-to-bam conversion, removal of PCR duplicates, recalibration, variant calling and annotation were described in the previous study[49]. Genomic VCF (g.vcf) files for each sample were generated by Genome Analysis Tool Kit (GATK) haplotype caller to ensure the coverage of non-variant sites[267].

Population control panel 1 (structure panel)

SNP-array data*:* We obtained array-based data from the publicly available populations from the Siberians and the Native Americans, which included 10 Siberian populations (Eskimo, Altai, Shors, Koryak, Chukchi, Teleuts, Evens, Yakut, Evenk, Buryat) from Cardona *et al.*[39]; 114 worldwide populations (including 15 Siberian populations and 51 New World populations) from Reich *et al.*[1]; 9 Sakha-Siberian populations from Fedorova *et al.*[268]; 13 Caucasus populations from Yunusbayev *et al.*[269]; 32 Turkic related populations (Siberians, Middle-Easterners, Central Asians and Caucasus) from Yunusbayev *et al.*[270]; 11 Siberians, Greenlanders and Athabascans from Rasmussen *et al.*[16]; 9 Altai region populations from Raghavan *et al.*[56] and 5 Siberian populations from Raghavan *et al.*[237] (**Table S1**).

Whole genome sequencing data: Whole genome sequencing data were also obtained from the 1000 genome project (1KGP) phase III, which included 26 populations[271], the Northeastern Siberian populations from Clemente *et al.*[40] and 9 current Old/New World populations[237]. We also

obtained ancient genomes from Siberia and the New World, which included a 4-kya Saqqaq genome[16]; a 24-kya Mal'ta Siberian genome[56]; a 11-kya Late Pleistocene human from Clovis[233]; 34 ancient Arctic people's genomes: including three from PreDorset culture, 14 from Middle Dorset, two from Late Dorset, two from Birnirk culture, one Norton and five Thules[9]; five Bronze age genomes from ancient Altai region[272]; the genome of the 8-kya Kennewick Man[234] and 23 ancient Native American genomes[237] (**Table S1**).

Population control panel 2 (selection panel)

European (CEU) and Han Chinese (CHB) populations from the 1KGP phase III[271] as well as the Northeast Siberian populations, including Chukchis, Eskimos, and Koryaks[40] were used in combination with the WES data from Nunavik Inuit to calculate the PBS.

Data merging and quality control

Merging genotype data: The genotype data of the aforementioned array-based population controls were merged with Nunavik Inuit using Plink 1.9[273] based on GRCh37/hg19 reference and dbSNP 131 database, with strand correlate with the 1KGP. SNPs without chromosomal positions and mitochondrial SNPs were removed. A total of 1,230,877 SNPs were included after the final merging, and were used as intervals for variant calling from the bam files of the ancient whole genome data aforementioned, as well as the 1KGP phase III data. After the removal of sex-chromosomal SNPs, the final merged data comprised of array-based contemporary populations, ancient populations and the 1KGP populations, which include 1,198,992 SNPs (1,198,070 SNVs) and a total of 5,422 individuals from 197 populations.

Quality control of the sequencing data: Variants from the g.VCF files of Nunavik Inuit were combined with the variants of CHB and CEU samples from the 1KGP phase III VCF files. High quality intervals of both datasets were generated and intersected with each other to provide high confident regions to obtain high quality exome variants. For the Nunavik Inuit WES data, high-confidence regions were considered to be that 80% individuals have 100% of reads in the refseq gene regions above 10X coverage, which contain 25,941,861 base pairs in the autosomal regions. The exome targets of the 1KGP phase 3 were also considered to be the high-confidence regions for this dataset, which contain 44,584,105 bp in the autosomal regions. GATK CombineVariants function was used to merge the datasets between the Nunavik Inuit (NUI) and the 1KGP phase3 CHB and CEU and the Northeast Siberians; and variant selection and merging were performed within the intervals of HC regions of both datasets. Indels and SNVs, which covered less than 10X in each population, were further removed in the combined dataset, non-polymorphic variants were also excluded.

A round of variant re-calling using GATK were performed on NUI, CHB and CEU BAM files using variants of the NUI-CHB-CEU merged dataset as intervals. This step was used for the calculation of population branch statistics (PBS) to ensure the accurate calling of homozygous reference genotypes in outgroup (CEU) and sister population (CHB).

Variant calling of WGS controls, including 25 North Siberians, 59 ancient genomes and 16 genomes from Old/New World populations, were also performed by GATK within the exome HC region described above. The VCF file was merged with four Native American populations (AMR) from the 1KGP and NUI-CHB-CEU variant recall file, using positions in NUI-CHB-CEU recall dataset as intervals.

Population Structure and Admixture

PCA and ADMIXTURE

Principle component analysis (PCA) implemented in Eigensoft 6.1[274] was performed on 1) a total of 5,422 individuals from the aforementioned structure panel (ALLPOP); 2) 4,611 individuals with the removal of African populations (NoAfr); and 3) 3,130 individuals selected from Siberian, Middle Asian, East European and Native American populations located along the migration path of the ancestors of today's Nunavik Inuit (NUIrelate). Each set has excluded the SNPs with the genotype missing rate over 0.5, 0.55 and 0.6, respectively; and LD-pruned to ensure that no pair of SNPs with r^2 greater than 0.5 in windows of 50 SNPs. The first 10 eigenvectors were calculated and maximum number of outlier removal iterations was set to 5 in the NoAfr and NUIrelate datasets.

The ancestry analysis was also performed on the three aforementioned datasets using ADMIXTURE software[231]. K=8 was used for ancestry estimation for each dataset based on pruned data used for PCA analysis. An additional PCA and ADMIXTURE analysis were also performed using the WES data of Nunavik Inuit and WGS controls to minimize ascertainment bias induced by SNP-chip.

PCAdmix[232] was used for the local ancestry estimation for individual genomes selected from highly admixed Inuit individuals and each of the populations with high level admixture of the Inuit ancestry. 12 populations were selected, including Naukan, Chikuchi, Eskimo, Yukagir, Koryak, Evens, Chipewyan, Cree, Ojibwa, Algonquin, Greenland Inuit and the Paleo-Eskimo. The data before applying PCAdmix analysis was phased using BEAGLE 4.1[275] with the 1KGP phase III as reference panel. We inferred three ancestral populations (ancient New World, European and

Inuit) based on the results from ADMIXTURE when preforming PCAdmix on highly admixed Inuit individuals. For each of the other 12 admixed individuals, three ancestral populations were inferred: for the Siberian populations (Naukan, Chikuchi, Eskimo, Yukagir, Koryak and Evens), the ancestries were inferred as Nunavik Inuit (NUI), Han Chinese (CHB) and Native American (NAT); for the Canadian-Arctic populations (Chipewyan, Cree, Ojibwa, Algonquin and Greenland Inuit), the ancestries were inferred as Nunavik Inuit (NUI), European (CEU) and Native American (NAT). The three ancestries for the Saqqaq individuals were ancient New World, Han and Inuit. The results were illustrated on the chromosomal ideographs.

LD decay and Runs of Homozygosity (ROH)

Genome-wide Linkage disequilibrium (LD) can be affected by genetic drift and natural selection; and admixture could also affect the length of LD thus causing faster LD decay. It is assumed that historical demography has an effect on LD in the isolated populations, such as Nunavik Inuit.

We calculated the D' from SNP-chip data of 14 indigenous and other present-day populations in order to compare the patterns of their LD decay. Each population contains 20-50 individuals to minimize the inflation. The five Nunavik villages (Inukjuak, Puvirnituq, Ivujivik, Salluit and Kangiqsualujjuaq) with more than 20 peoples were selected; along with three Northeast Siberian populations including Naukan, Eskimo and Chukchi; Athabaskans from Alaska; Chipewyans from NWT in Canada; West/East Greenlanders; CHB and CEU populations.

LD measured by D is

$$D=Frequency(AB)-Frequency(A)Frequency(B)$$

$$Or\ D=Frequency(AB)Frequency(ab)-Frequency(Ab)Frequency(aB)$$

The AB alleles are in LD if $D \neq 0$

And for the calculation of D'

If $D>0$,

$$D' = D/D_{max}(\min(Frequency(A)\ Frequency(b),\ Frequency(a)\ Frequency(B)))$$

If $D<0$,

$$D' = D/D_{min}(\max(-Frequency(A)\ Frequency(B),\ -Frequency(a)\ Frequency(b)))$$

The result of the LD decay for each population was plotted in R.

Runs of Homozygosity (ROHs) were used in estimating individual genome-wide autozygosity, with 5,000 kb of sliding window and homozygous SNPs spanning 1,000 kb used to define the homozygous segment. The SNP-chip data of Chukchi, Naukan, Eskimo from Siberia, Athabascan and Chipewyan from Alaska and Canada, East and West Greenlandic populations and YRI, CHB and CEU populations from 1KGP were compared with the Nunavik individuals from different villages. The proportion of each population with one or more ROH up to 1.5 Mb, between 1.5-2.49 Mb, 2.5-4.99 Mb, 5.0-9.99 Mb and more than 10 Mb in length was calculated respectively.

Treemix

We used Treemix[235] to infer population splits and mixtures from genotypes of SNP-chip data in selected populations that closely related to Nunavik Inuit (NUIrelate). The algorithm used Gaussian approximation to estimate genetic drift between populations. Thirty-four ancient Arctic people's genomes and 23 ancient Native American genomes were excluded from the tree

generation because of their low sequencing quality. Different populations focusing on the Old/New world peoples were included in the constructing of trees with or without high quality ancient genomes (Saqqaq, Mal'ta, Clovis, Kennewick and four ancient Bronze Age Altai peoples) because they were shown as outliers. 50 trees were generated for each dataset with 0-10 migration events, the trees with the highest likelihood were presented.

Detection of Selection by Population Branch Statistics (PBS)

PBS

Population-wide site variant frequencies were calculated from the NUI-CEU-CHB exome HC variant dataset described above. Variants with no frequency in one population was assigned a minimum MAF to reduce the bias introduced by a limited population size. Fixation index (F_{ST}) value for each variant was calculated between each of the two populations, and the classical transformation was used as follows,

$$T = -\log (1 - F_{ST})$$

The value T was calculated between the Inuit and CEU populations (T^{IE}), the Inuit and CHB populations (T^{IH}) and CEU and CHB populations (T^{EH}). PBS values, representing the lengths of the branch leading to the Nunavik Inuit population since its divergence with Han Chinese, were calculated as

$$PBS^{I} = (T^{IE} + T^{IH} - T^{EH})/ 2$$

when using Europeans as outgroup[236].

The results were shown in Manhattan plot and all variants with PBS > 0.1 were plotted against concatenated chromosomal locations. The exome variants with PBS > 0.1 were shown in red. We also calculated the PBS values between Nunavik Inuit and Northeastern Siberians, with CHB as outgroup to identify more recent selections.

Expression analyses of the genes under selection

Genes under potential natural selection were chosen for real-time quantitative PCR (RT-qPCR) analysis to detect whether their levels of expression are significantly changed among the Nunavik Inuit samples. EBV-transformed lymphoblastoid cell lines (LCL) were selected from 15 unrelated Nunavik Inuit and 14 unrelated French Canadians (FCs) as controls. RNA was extracted from each LCL and cDNA was prepared using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). RT-qPCR was performed using the TaqMan Gene Expression Assay on QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems). For each assay, RNA polymerase II polypeptide A (POLR2A) was used as endogenous control to normalize the gene expression level. Three analyses were done with independent RNA extractions and each sample was tested in triplicate. The triplicates were valid if the delta Ct standard deviation was smaller than 0.25, and the mean delta CT value over three replications was used and with delta delta CT calculated using one FC individual as the calibrator. Relative quantification (RQ)[276] for each gene was calculated for 14 Nunavik and 14 FC individuals using the formula of:

$$RQ = 2^{-\Delta\Delta Ct}$$

And were compared by non-parametric Mann-Whitney U test with the RQ of each gene subsequently plotted.

Susceptibility of IA in Nunavik Inuit

Among the Inuit individuals recruited in this study, forty-two of them were also diagnosed to have definite IA, a cerebrovascular disorder affects the brain vascular walls (thinning and bulging) therefore leading to subarachnoid hemorrhage[277]. We sought to discover if the gene(s) that have undergone natural selection in Nunavik Inuit would potentially influence their susceptibility of IA, because the changes of life style of today's Nunavik Inuit may have possibly introduced new selective pressure on specific genetic variants that they have carried over through their population history.

Gene ontology (GO) enrichment analyses were performed on a number of genes under selection (PBS>0.3) between Nunavik Inuit and CHB using Panther (http://geneontology.org/). SNPs with batch bias (<0.05 after multiple correction) in Nunavik Inuit dataset were removed; and we performed association tests in 42 un-admixed Nunavik IA cases and 62 un-admixed Nunavik controls focused on the high quality, potential protein altering variants which are under strong selection (PBS>0.8). Bonferroni correction was applied to adjust the results with $P < 0.05$ as significance.

Acknowledgements

We thank the Nunavik Inuit participants, the Nunavik communities and the clinicians for their support and contributions of this research.

Figures

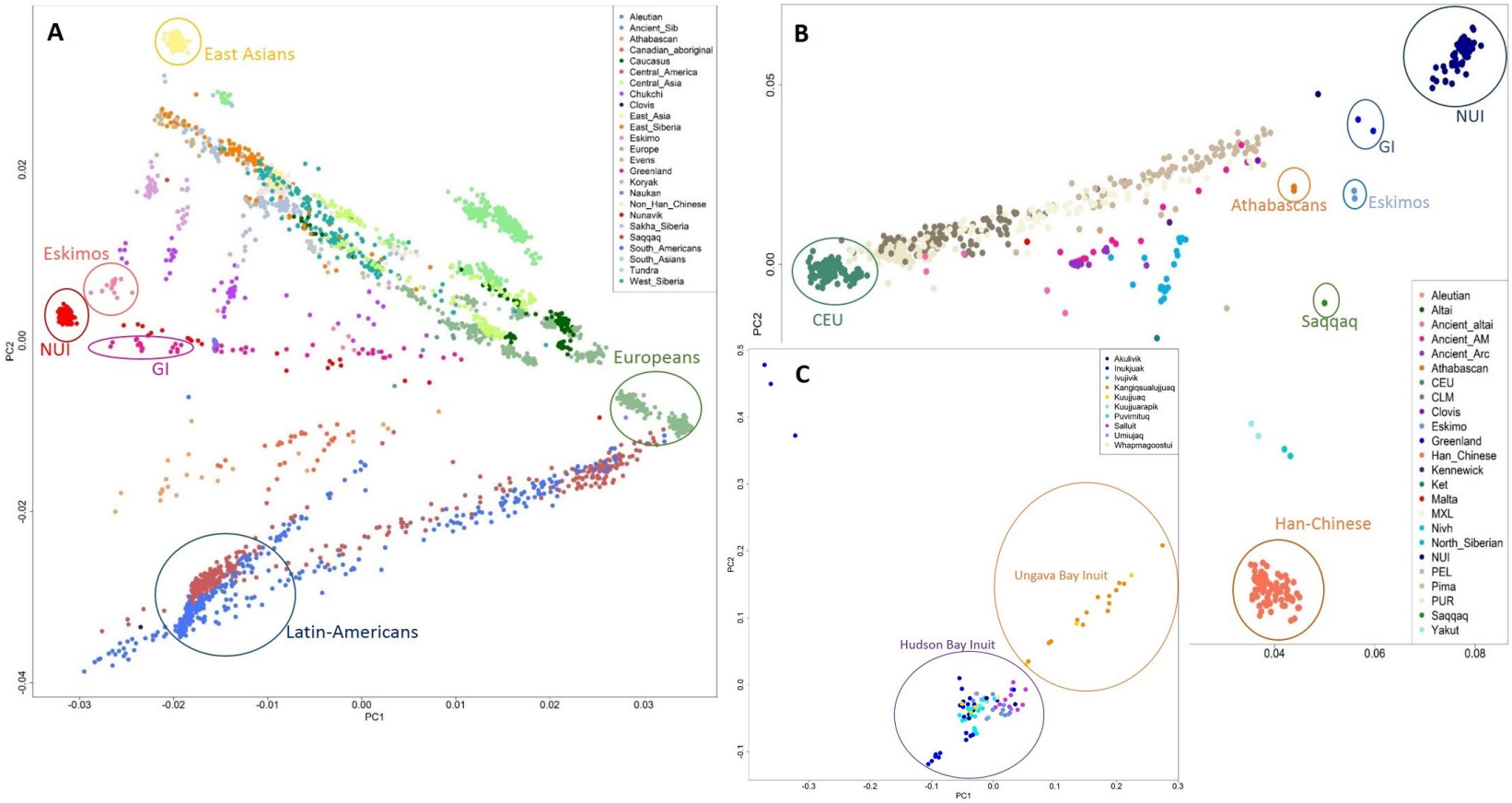

Figure 1. Principal component analysis (PCA) of Nunavik Inuit and other populations.

A. PCA using SNP genotyping panel "NUIrelate" showing the Nunavik Inuit and selected indigenous populations. B. PCA using WES data showing the Nunavik Inuit and selected Arctic and Native American indigenous populations. C. PCA using WES data showing the Inuit from the Hudson Bay and the Ungava Bay of Nunavik.

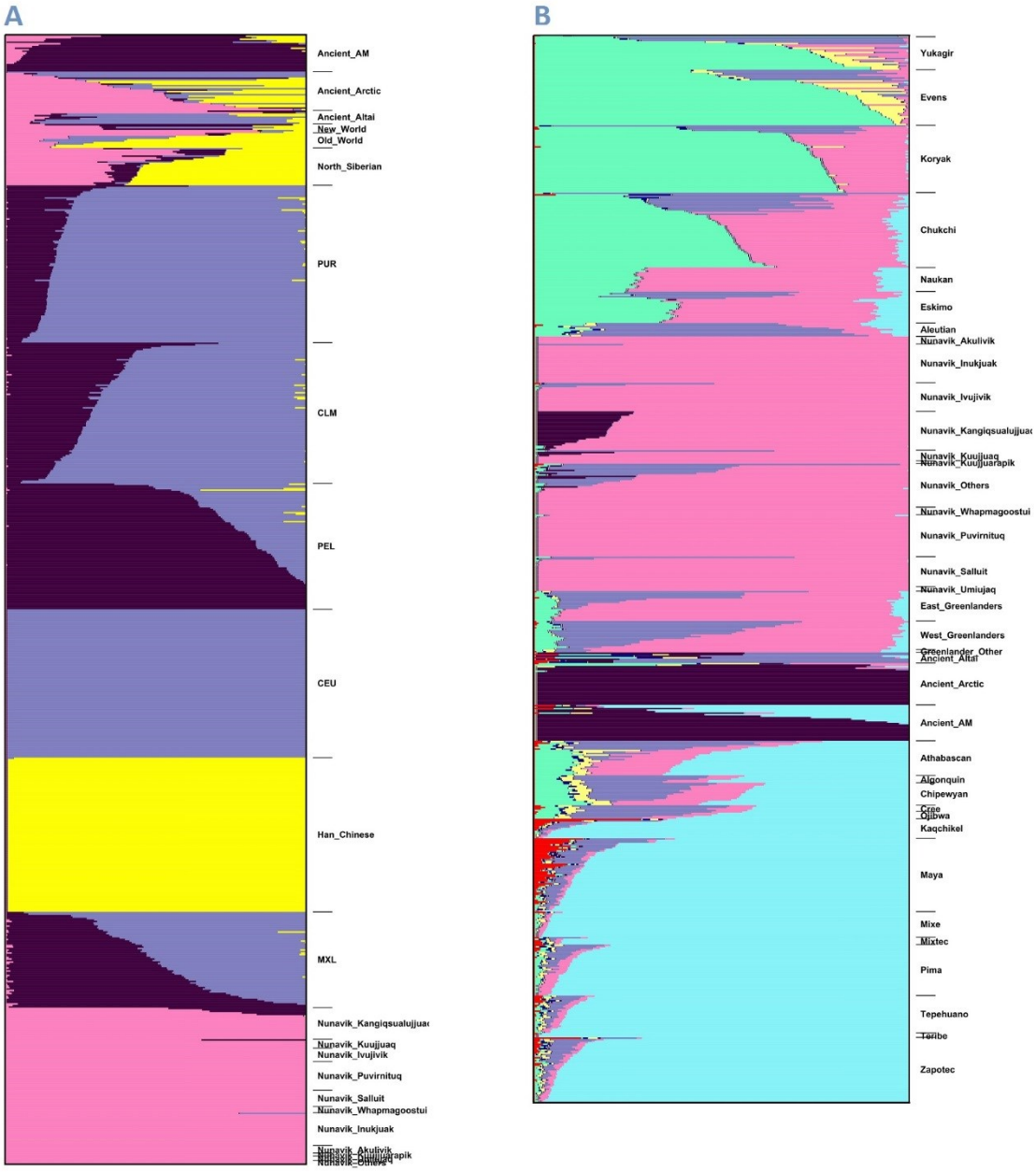

Figure 2. Estimated admixture proportions for Nunavik Inuit with Arctic and Native American indigenous populations.

The ancestry components are estimated with ADMIXTURE using A. the WES data of the Nunavik Inuit (by villages) and 1KGP Native American populations and the ancient populations (K=4); and B. the SNP-chip genotyping data, selected populations with Inuit ancestry are shown (K=8).

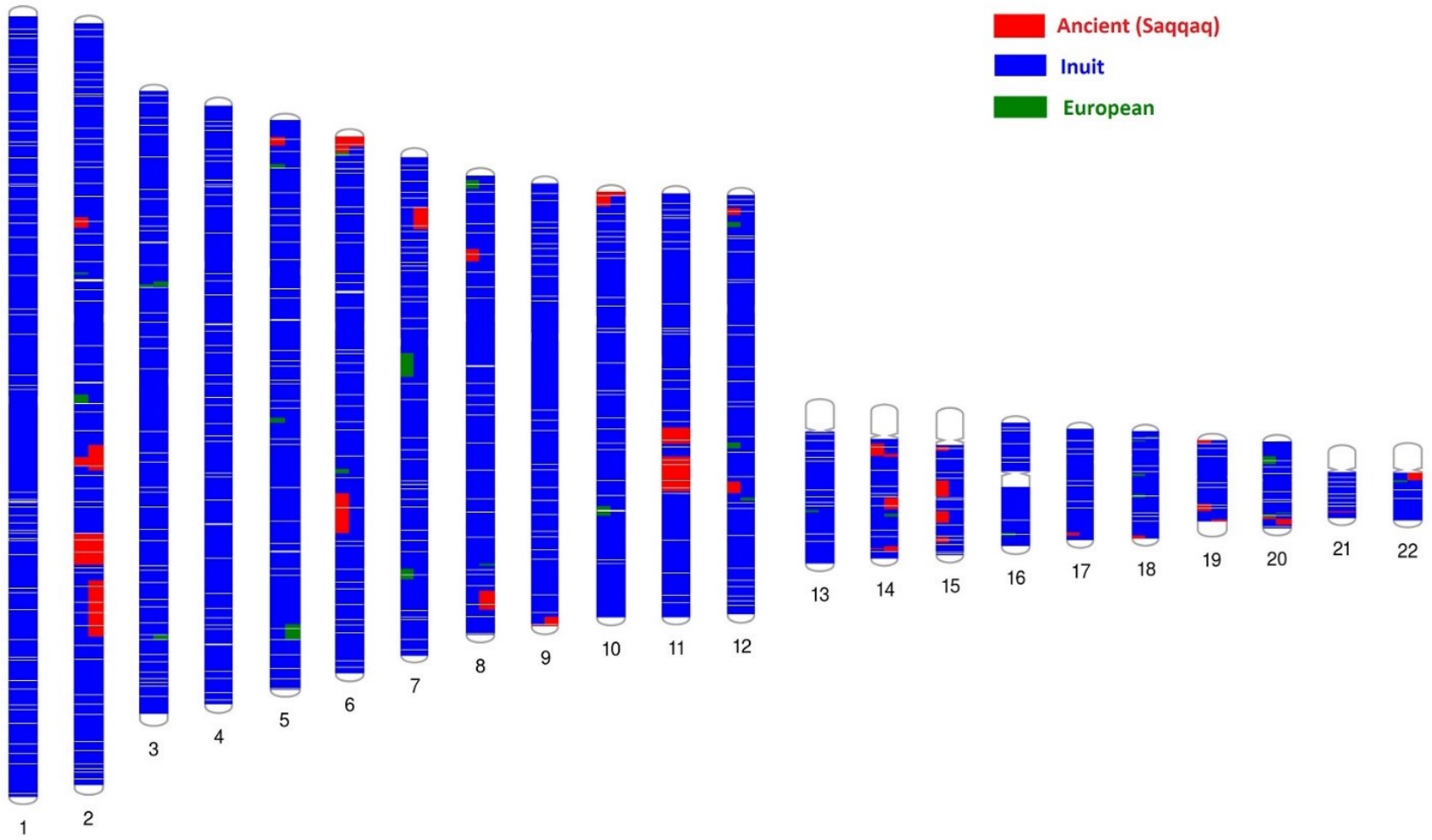

Figure 3. Local ancestry inference of an individual from Kangiqsualujjuaq.

The individual genome potentially come from the Saqqaq ancestry are shown in red (3.8%).

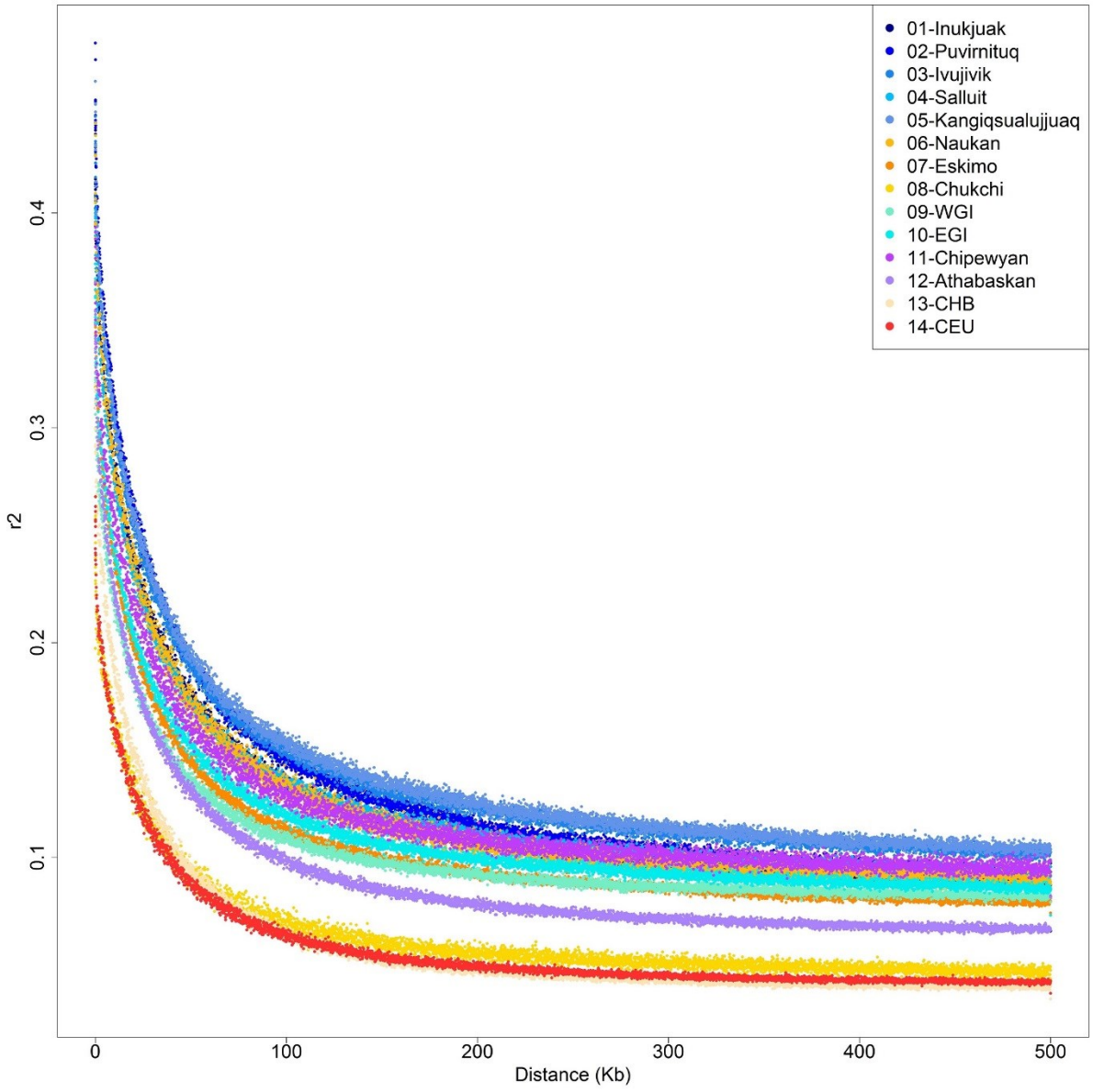

Figure 4. LD decay for different populations.

The figure depicted the LD decay for Nunavik Inuit villages (01-05), Northeastern Siberians (06-08), Greenlandic Inuit (09-10), Canadian Arctic indigenous peoples (11-12) and two reference populations, CHB and CEU from 1KGP (13-14).

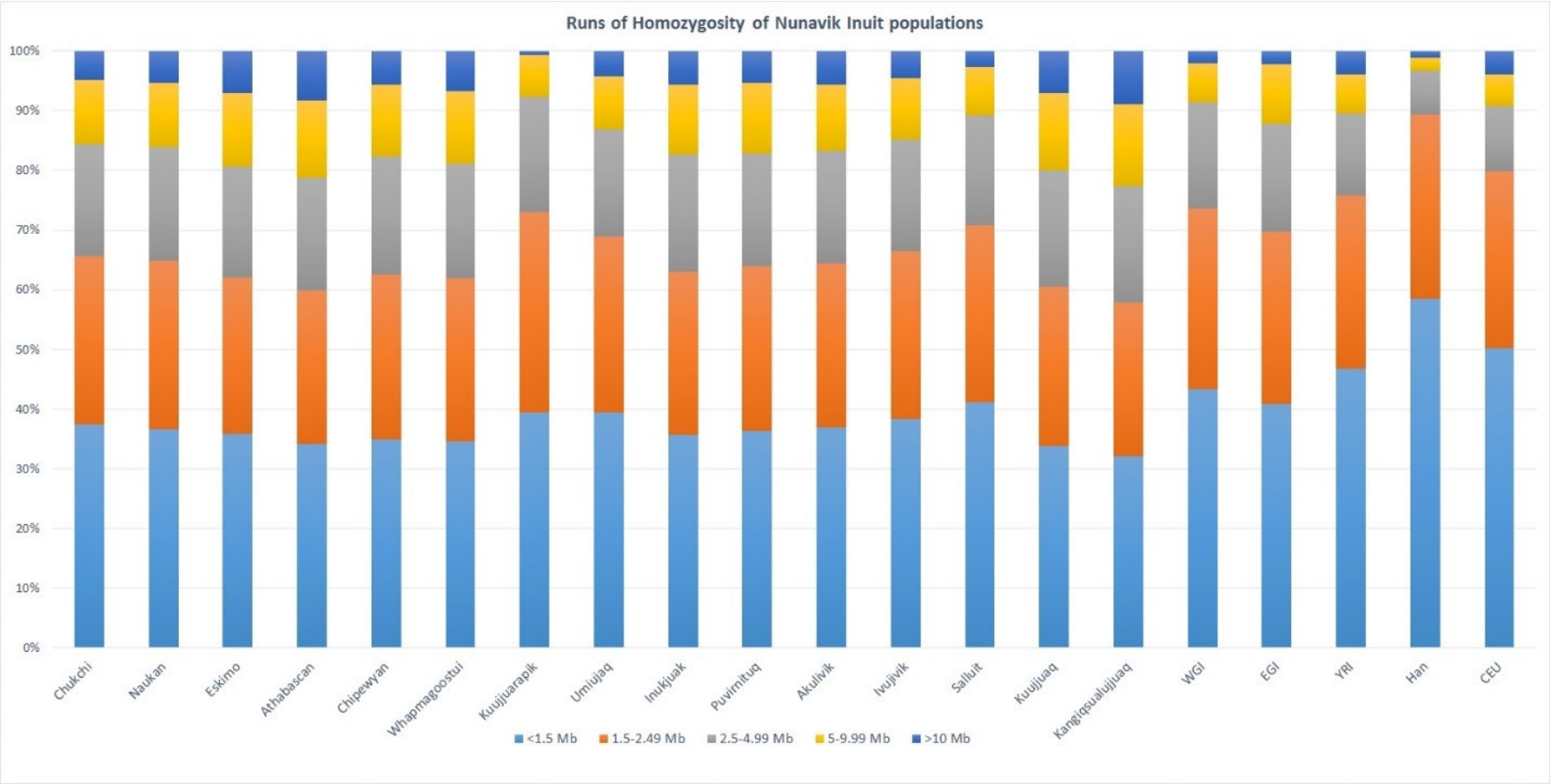

Figure 5. Estimated Runs of Homozygosity (ROH) lengths of different populations.

Including Nunavik Inuit from different villages, Arctic indigenous peoples and 1KGP populations. ROH in windows of different sizes were shown in proportion in each population.

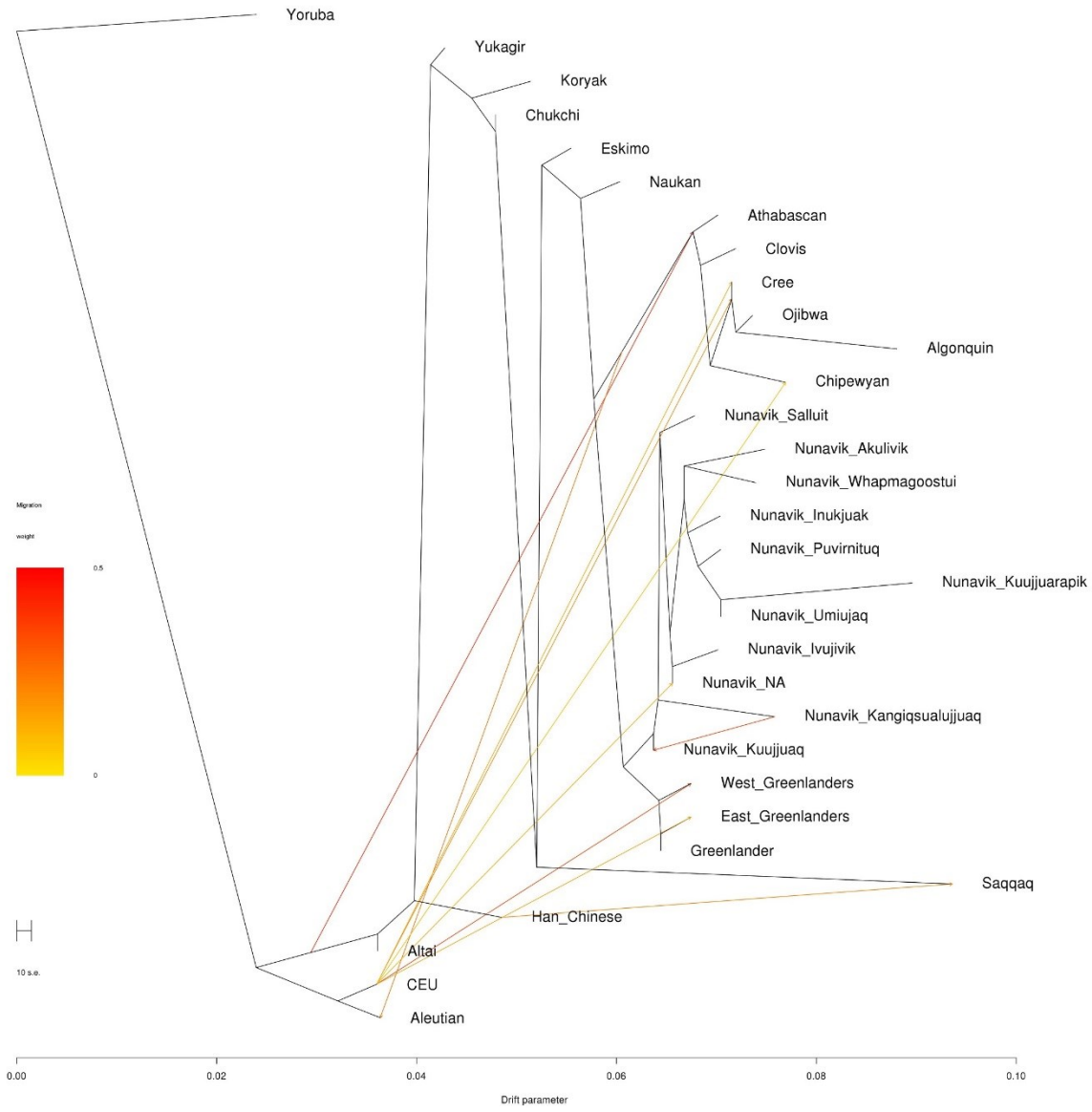

Figure 6. Maximum-likelihood genetic-drift tree topology relating people from Nunavik Inuit and other Arctic indigenous populations.

Estimated by Treemix, with 10 migration events.

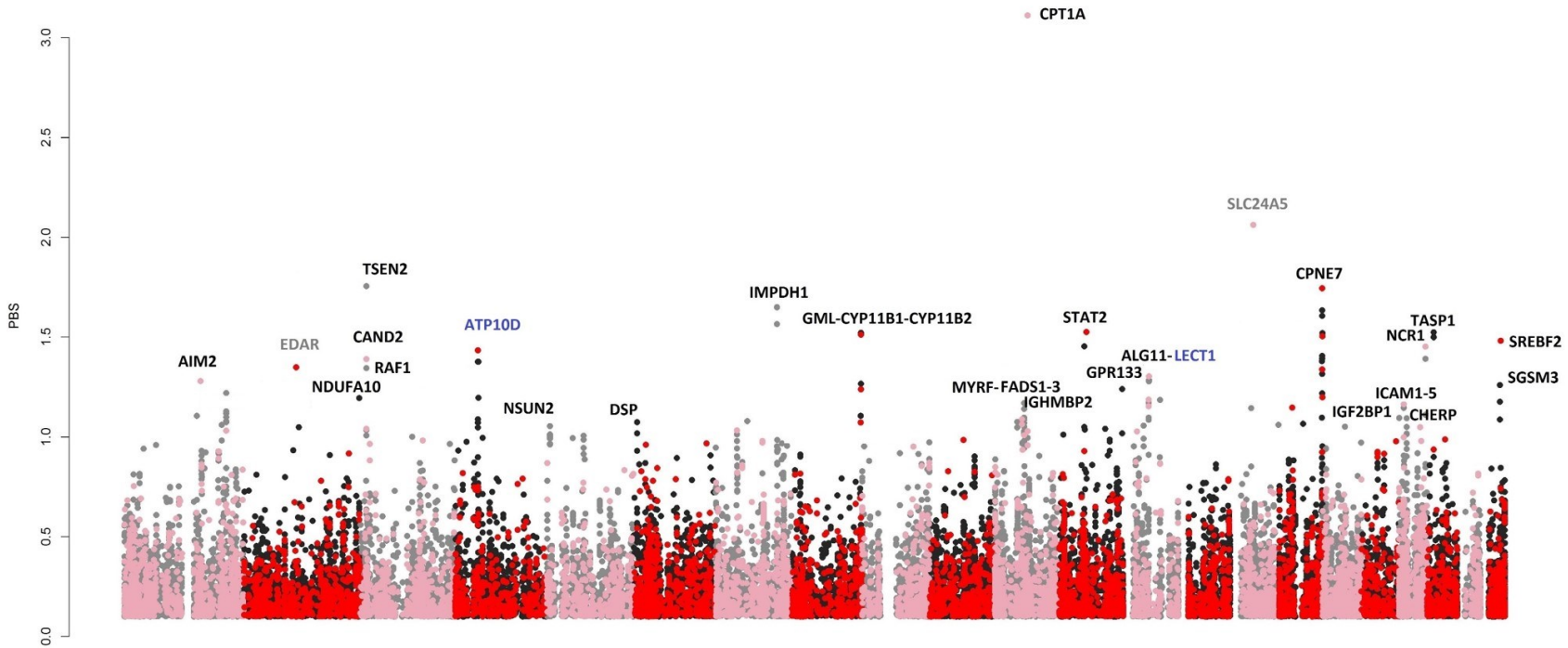

Figure 7. Manhattan plot displaying Population Branch Statistics (PBS) of Nunavik Inuit.

This figure shows the F_{ST} based PBS results, which measures the variant frequency changes of each site. CHB is used as the sister population and CEU is used as outgroup, variants with PBS>0.1 are displayed, with red (pink) dot represent exonic variants. 30 genes marked on the figure were also included in the subsequent expression analysis.

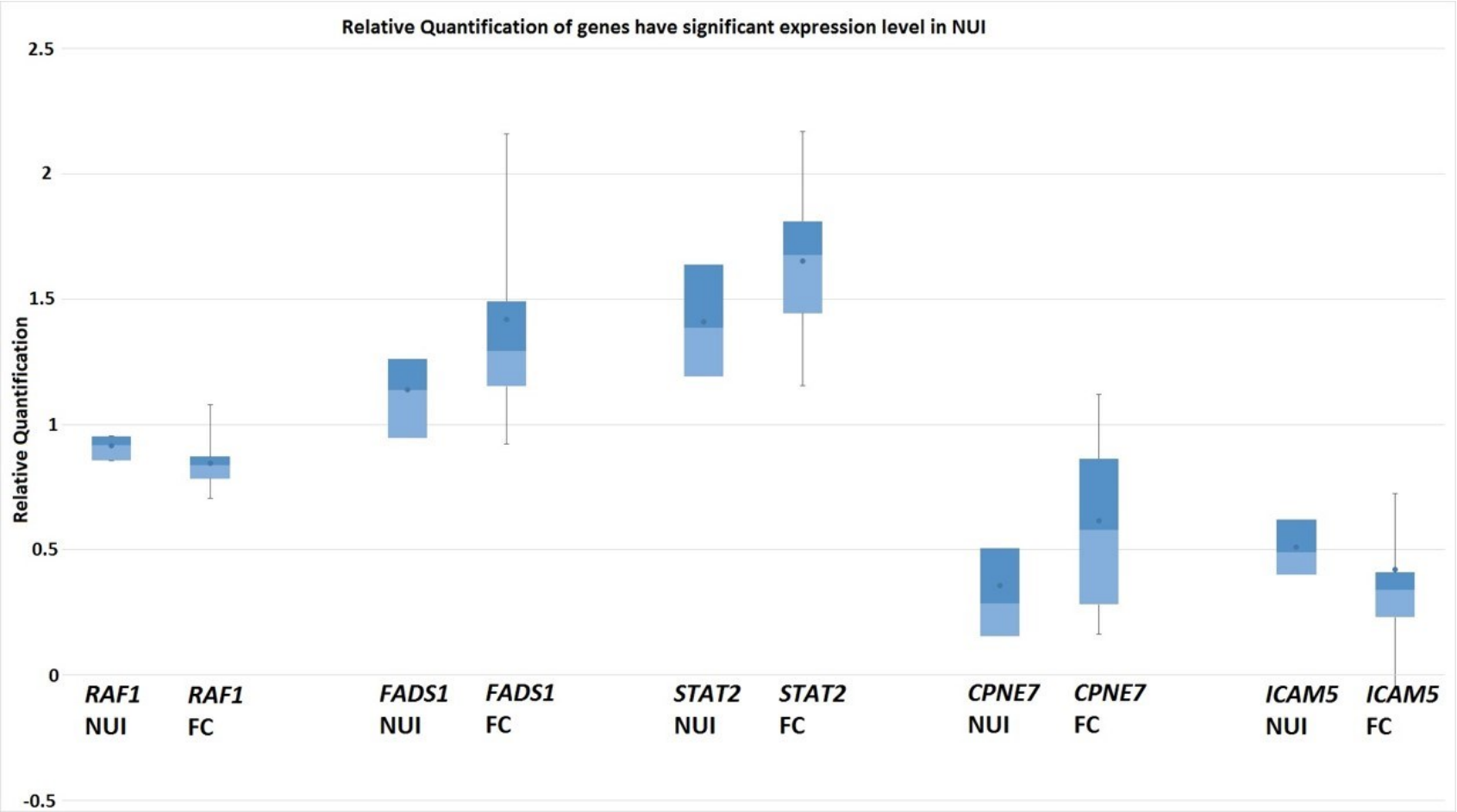

Figure 8. The relative quantification (RQ) scores of genes with significant expression level changes between Nunavik Inuit and FC.

Tables

Table I. Candidate genes under positive selection based on top PBS scores in Nunavik Inuit using WES data

Gene	Top SNP	Top PBS in NUI-CHB-CEU	Function(CADD)	MAF.NUI	MAF.CHB	MAF.CEU	Top PBS in NUI-NES-CHB	Top SNP (PBS>1) in GI-CEU-CHB	No. SNPs with PBS>1 in NUI
CPT1A	rs80356779	3.11191	missense(18.21)	0.9567308	0	0	0.839809	NA	1
TSEN2	rs735640	1.75609	3'-UTR(NA)	0.8300971	0	0	NA	rs735640 (1.386817257)	1
CPNE7	rs12445560	1.74435	near-splice(9.53)	0.9471154	0.126214	0.09596	0.48916	rs139901937 (1.386817257)	4
IMPDH1	rs72624969	1.65097	intronic(NA)	0.9759615	0.286408	0.07071	NA	rs4731447 (1.074576366)	3
STAT2	rs2066815	1.526	missense-near-splice(25)	0.7932692	0.009709	0	0.839179	NA	1
TASP1	rs11697393	1.52427	intronic(NA)	0.8300971	0.048544	0.0202	NA	rs11697393 (1.145574318)	3
CYP11B1	rs4534	1.51184	missense(0.09)	0.9615385	0.446602	0.01515	0.495403	rs57589970 (1.42684416116379)	1
SREBF2	rs2228314	1.48112	missense	0.9951923	0.194175	0.25253	NA	rs2228314 (1.052002479)	1
NCR1	rs2278428	1.45223	missense(5.939)	0.9567308	0.320388	0.06566	0.510609	rs2278427(0.801)	3
ATP10D	rs16851681	1.43361	missense(0.09)	0.9807692	0.194175	0.23232	1.073339	rs13152689 (1.01525329436256)	5
CAND2	rs180768267	1.38952	missense(16.91)	0.7884615	0.021739	0	1.094545	rs181307051 (1.067382254)	2
RAF1	rs5746223	1.34431	intronic(NA)	0.8267327	0.033981	0.09596	NA	rs5746223(0.72)	3
LECT1	rs62637607	1.28517	intronic(NA)	0.7355769	0	0.0101	NA	rs62637607 (1.376472464)	1
AIM2	rs2276405	1.27947	missense(16.34)	0.75	0.029126	0	0.634696	rs2276405 (2.536097555)	1
CYP11B2	rs6432	1.26684	intronic(NA)	0.9509804	0.519608	0.01515	NA	rs4536 (1.30937734894169)	1
SGSM3	rs55844816	1.25915	intronic(NA)	0.8798077	0.169903	0.09596	NA	rs55844816 (1.413484817)	4
GPR133	rs1212936	1.24033	intronic(NA)	0.9663462	0.160194	0.32828	NA	rs1195923(0.671033087505734)	2
GML	rs3764795	1.23792	missense(15.17)	0.038835	0.504854	0.95455	0.368187	rs3750247(0.238465870959858)	1
NDUFA10	rs77816205	1.19479	intronic(NA)	0.7548077	0	0.06633	NA	rs77816205 (1.427053791)	1
ALG11	rs17480245	1.18738	missense(0.017)	0.7058824	0	0.0101	0.532059	rs17480245 (1.212805669)	1
FADS1	rs174556	1.16915	intronic(NA)	0.0048077	0.664706	0.68966	NA	rs174547(0.763990036106579)	3

ICAM5	rs1056538	1.16354	missense(16.96)	0.0192308	0.81068	0.61616	1.038068	rs2228615(0.546499505209769)	3
FADS2	rs174602	1.14054	intronic(NA)	0.9326923	0.257282	0.19192	NA	rs174602 (1.14510508)	1
CHERP	rs12460141	1.09348	intronic(NA)	0.6923077	0.005952	0.00549	NA	rs12460141 (1.508902039)	2
MYRF	rs174536	1.08532	intronic(NA)	0.0096154	0.674757	0.67677	NA	rs108499(0.770196821564124)	4
DSP	rs7741957	1.07435	intronic(NA)	0.8543689	0.315534	0	NA	rs7741957 (1.907001643)	2
NSUN2	rs506416	1.05426	intronic(NA)	0.0148515	0.650485	0.70202	NA	rs6887702(0.781014952248378)	4
IGF2BP1	rs4265867	1.05156	intronic(NA)	0.6634615	0	0	NA	rs4265867 (1.735874019)	1
IGHMBP2	rs560096	1.02829	missense(1.24)	0.0480769	0.538835	0.88384	0.15293	rs1249463(0.954137807645166)	1
ICAM1	rs5498	0.99744	missense(0.754)	0.0192308	0.737864	0.58586	0.881544	rs5498(0.424053850741787)	1

2.3: INCREASED MISSENSE MUTATION BURDEN OF FATTY ACID METABOLISM RELATED GENES IN NUNAVIK INUIT POPULATION

Manuscript: Increased missense mutation burden of fatty acid metabolism related genes in Nunavik Inuit population

PLoS One, 2015 May 26; 10(5):e0128255. doi: 10.1371/journal.pone.0128255

Authors and affiliations:

Sirui Zhou[1,2], Lan Xiong[2,3], Pingxing Xie[1,4], Amirthagowri Ambalavanan[1,4], Cynthia Bourassa[1], Alexandre Dionne-Laporte[1], Dan Spiegelman[1], Maude Turcotte Gauthier[1], Edouard Henrion[1], Ousmane Diallo[1], Patrick A. Dion[1,5], Guy A. Rouleau[1,5,*]

1) Montreal Neurological Institute and Hospital, McGill University, Montréal (Que), Canada; 2) Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal (Que), Canada; 3) Centre de recherche, Institut universitaire en santé mentale de Montréal (Que) Canada; 4) Department of Human Genetics, McGill University, Montréal (Que) Canada; 5) Department of Neurology and Neurosurgery, McGill University, Montréal (Que) Canada.

Authors contribution:

Study design: SZ, LX, GAR, PAD

Experiment: SZ, CB

Bioinformatic analysis: SZ, AA, PX, ADL, DS, EH, OD

Sample recruitment: MTG

Manuscript writing: SZ, LX, PAD, GAR

Abstract:

Background: Nunavik Inuit (northern Quebec, Canada) reside along the arctic coastline where for generations their daily energy intake has mainly been derived from animal fat. Given this particular diet it has been hypothesized that natural selection would lead to population specific allele frequency differences and unique variants in genes related to fatty acid metabolism. A group of genes, namely *CPT1A, CPT1B, CPT1C, CPT2*, *CRAT* and *CROT*, encode for three carnitine acyltransferases that are important for the oxidation of fatty acids, a critical step in their metabolism. **Methods:** Exome sequencing and SNP array genotyping were used to examine the genetic variations in the six genes encoding for the carnitine acyltransferases in 113 Nunavik Inuit individuals. **Results:** Altogether ten missense variants were found in genes *CPT1A, CPT1B, CPT1C, CPT2* and *CRAT*, including three novel variants and one Inuit specific variant *CPT1A* p.P479L (rs80356779). The latter has the highest frequency (0.955) compared to other Inuit populations. We found that by comparison to Asians or Europeans, the Nunavik Inuit have an increased mutation burden in *CPT1A, CPT2* and *CRAT*; there is also a high level of population differentiation based on carnitine acyltransferase gene variations between Nunavik Inuit and Asians. **Conclusion:** The increased number and frequency of deleterious variants in these fatty acid metabolism genes in Nunavik Inuit may be the result of genetic adaptation to their diet and/or the extremely cold climate. In addition, the identification of these variants may help to understand some of the specific health risks of Nunavik Inuit.

Introduction:

Modern Inuit are descendants of the Dorset peoples from the second wave of migrations which took place over 1,000 years ago[278]. Previous studies have established that these Inuit are genetically closest to the contemporary north-east Siberian populations[16], and belong to one of the three native ancestral groups that populated the Americas. The ancestors of today's Nunavik (a region of northern Quebec) Inuit had migrated from Alaska to the east across the far north. With a population of barely 10,000 individuals, Nunavik Inuit form a very isolated population whose ancestral genetic profiles are likely to be well preserved until today. The Inuit enrolled in this study were recruited from 13 of the 14 inhabited villages, and they represent over 1% of the Nunavik Inuit population. As a part of their unique lifestyle, Nunavik Inuit derive approximately 75% of their daily energy intake from animal fat[279]. The traditional diet (high in fat and low in carbohydrates) of Nunavik Inuit suggests that they have high rates of gluconeogenesis, which is supported by their larger liver size compared to other populations.

The oxidative processing of fatty acids is critical for generating energy from a diet enriched with animal fat. Several enzymes are involved in the metabolism of fatty acids, including the carnitine acyltransferases. The carnitine acyltransferase gene family is comprised of six members which encode for three types of enzymes: (1) carnitine palmitoyltransferases (CPTs) encoded by *CPT1A, CPT1B, CPT1C* and *CPT2*; (2) carnitine acetyltransferase (CrAT) encoded by *CRAT*; and (3) carnitine octanoyltransferase (CrOT) encoded by *CROT*[280]. The genomic features of these six genes are listed in **Table II.** CPTs include two members, CPT-I and CPT-II. As a rate-limiting step in fatty acid oxidation, CPT-I converts long-chain fatty acyl molecules into their

corresponding acylcarnitines, which are then transported across the inner mitochondrial membrane for beta-oxidation. CPT-II is responsible for the reversal of this process[281]. CrOT and CrAT catalyze the reversible transfer of fatty acyl groups between CoA and carnitine, and therefore determine the acetyl-CoA/CoA ratio[282].

CPT-I and II deficiencies are genetic disorders that are characterized by the reduced CPT enzyme activity which can lead to severe consequences, such as liver failure and renal problems[283; 284]. Although these deficiencies are commonly caused by autosomal recessive mutations in *CPT1A* and *CPT2*, heterozygous carriers of *CPTs* mutations can also be symptomatic[285].

Considering their critical roles in lipid metabolism and energy generation, the six genes encoding carnitine acyltransferases may be subjected to evolutionary pressure, especially in populations such as the Inuit, which require an increased enzyme activity because of their special diet and extreme living conditions. Recent studies have shown that the diet of Nunavik Inuit led to changes in specific biomarkers. For instance, Nunavik Inuit have very high levels of n-3 polyunsaturated fatty acids (PUFA), especially in the elderly[286], as well as high levels of plasma LDL cholesterol[287]. It is possible that these changes may be partially due to the variations in the carnitine acyltransferase genes, which can affect their enzymatic activity. Because of these biochemical characteristics, it is important to study the genes that are critical to fatty acid metabolism. In addition to providing insight about the Nunavik Inuit adaptation, the identification of variants in these genes may also help to understand some of their specific health risks. To examine this hypothesis, we used exome sequencing and SNP array genotyping to investigate genetic variants across these six genes in a cohort of Inuit from Nunavik.

Materials and Methods:

Subjects

We recruited 113 Inuit individuals (62 males and 51 females, mean age 52 years) from 13 inhabited villages of Nunavik. The study was approved by Comité d'éthique de la recherche du Centre hospitalier de l'Université de Montréal (ND 04.101) (Québec, Canada) and the Nunavik Nutrition and Health Committee (Québec, Canada). Individual written consent was obtained from each participant before entering this study.

Exome sequencing and coding variant identification from six carnitine acyltransferase genes

The genomic DNA of 113 Nunavik Inuit individuals was extracted from peripheral blood lymphocytes using Gentra Systems PUREGENE DNA purification kit (Qiagen). A 50 µl DNA sample at a concentration of 100 ng/µl from each sample was captured by Agilent SureSelect 50mb/V4 capture kit. The library was subsequently sequenced at 100 bp pair-end using Illumina Hiseq 2000, with 3 samples per lane to ensure an average coverage depth of 100-fold.

Raw fastq files were aligned to NCBI human reference GRCh37 using Burrows-Wheeler Aligner (BWA)[288], with all PCR duplicates removed from the alignments. The aligned reads were converted to binary format for further analysis using Sequence Alignment/Map (SAM) tools[288]. Single nucleotide variant (SNV) and insertion/deletion (indel) calling were performed using Genome Analysis Toolkit (GATK) version 2.7[267]. Variant annotation was performed using ANNOVAR program[289] with references to GRCh37/hg19, dbSNP version 132, 1000 Genomes project (1KGP) (2012 data release)[290], 69 Complete Genomics (2012 update) and exome variant server (EVS) with approximately 6,500 exomes (NHLBI-ESP project, 2013 update)[291]. Finally,

variant segregation analysis was performed using an in-house segregation program, using more than 1,000 exomes of different ethnicities from our lab as controls.

In the variant analysis, intronic, intergenic and non-coding variants were excluded because of insufficient coverage. Quality filters were set at sequencing depth ≥20x and variant frequency ≥25%, with genotype quality ≥10. Rare variants were defined as variants with minor allele frequency (MAF) ≤0.01 in the aforementioned public databases. All potential missense and splicing site variants from the genes *CPT1A, CPT1B, CPT1C, CPT2*, *CRAT* and *CROT* were examined, and were further validated by manual inspection using the Integrative Genomics Viewer 1.4 (IGV)[292] and Sanger sequencing.

SNP genotyping and common variant identification from six carnitine acyltransferase genes

SNP genotyping was performed on 113 Nunavik Inuit, using Illumina HumanOmniExpress-12 SNP array, which contains 730,525 SNPs. Raw data was processed and analyzed using Illumina GenomeStudio and quality control was performed by PLINK 1.07[293]. SNPs within each gene region and 2,000 bp up/downstream flanking regions were extracted from all individuals. Variant concordance within these regions was calculated between data from exome sequencing and SNP array genotyping. Allele frequencies and Hardy–Weinberg equilibrium (HWE) of all selected SNP were calculated using PLINK. Calculation of linkage disequilibrium (LD) and haplotype analysis were performed using Haploview 4.2[294].

Statistical analysis

Multidimensional scaling (MDS) and Admixture analyses[231] were performed on the Nunavik Inuit genotype data against HapMap genotype data from Asian (CHB and JPT), European (CEU) and African (YRI) populations to assess the mixture of Nunavik Inuit population.

Missense variations of the carnitine acyltransferase genes were retrieved from 286 Chinese (CHB) and Japanese (JPT) from the 1KGP database and European descendants from EVS, to serve as controls.

Mutation burden is defined as the average number of rare (MAF $\leq$0.01) missense variant alleles per person. To examine whether there is an increased mutation burden in the selected genes in Nunavik Inuit compared to other populations, we performed a binomial test to compare rare allele frequencies between Nunavik Inuit and 1KGP Asians. The Adaptive Permutation test was also performed to compare the mutation burden in Nunavik Inuit with Asians. The Permutation test was performed five times based on the number of variant-containing genes, with *Bonferroni* correction for multiple comparisons applied and $p < 0.01$ was set as statistically significant.

We also calculated two F-statistics (F_{ST} and F_{IS}) scores based on variants identified in the carnitine acyltransferase genes. The F_{ST} value provides an indication about the population genetic difference that is due to the genetic drift between subpopulations. The F_{IS} value measures the proportion of total inbreeding within a population that is due to non-random mating within subpopulations[295]. To further compare the impact of these variants in Nunavik Inuit to other populations, we generated a scatterplot using the R package "ggplot2" of all functional variants identified in *CPT1A, CPT1B, CPT1C, CPT2* and *CRAT* in Nunavik Inuit, 1KGP Asians and EVS Europeans. Mann-Whitney U test was performed to compare the PolyPhen-2 scores of all rare missense variants found in Nunavik Inuit and 1KGP Asian populations, with $p<0.05$ as significant.

Results:

Data quality control

Using the data from the HumanOmniExpress SNP array, four individuals with sex mismatch were identified by PLINK and removed from the subsequent analysis. After MDS testing, the genomes of nine additional individuals were found to have a mixture of both Inuit and European genomes and were therefore also removed from subsequent analysis. This mixture is probably the results of intermarriages between Nunavik Inuit and French-Canadians. Therefore a total of 100 Inuit samples were included in the genetic/statistical analysis (**Figure 9**).

Based on GATK's DepthOfCoverage Walker[267] program, the average coverage depth of the coding regions of *CPT1A, CPT1B, CPT1C, CPT2, CRAT* and *CROT* were 102X, 134X, 107X, 102X, 101X and 80X, respectively. The percentage of targeted region with coverage depth of more than 20X for each gene were 92.7%, 97.3%, 99.6%, 94.1%, 98.6% and 90.2%, respectively (**Figure S1**).

Of the 6 carnitine acyltransferase genes, coding variants were identified only in *CPT1A, CPT1B, CPT1C, CPT2* and *CRAT* in Nunavik Inuit. Since no coding variants were found in *CROT*, all subsequent analyses and discussions were focused on the other 5 genes. Across these genes, 16 variants were identified by both the exome sequencing and SNP array, 4 of which were exonic variants. The concordance rate of the exonic variants between the two platforms was 100%, while for the intronic variants it was 77% (**Table S1**). VCF (Variant Call Format) files of the exome sequencing data and PLINK genotype files of the SNP array covering the carnitine acyltransferase gene regions can be found in supporting information-Compressed/ZIP File Archive.

Missense variants of the carnitine acyltransferase genes in different populations

In the 100 individuals included in the analysis, 10 missense and 3 synonymous variants were found in 5 carnitine acyltransferase genes. Three missense variants, *CPT1C* p.T265M, *CPT2* p.R477W, *CRAT* p.S78F and one synonymous variant *CPT1A* p.V616V, were unique to the Nunavik Inuit population (**Table III** and **Table S2**). In addition, an Inuit specific variant *CPT1A* p.P479L was also observed. These five variants and a rare synonymous *CRAT* p.A575A variant were validated by Sanger sequencing (**Figure S2**). The frequencies of the 10 missense variants were compared between the Nunavik Inuit, Asian, European and African populations. *In silico* prediction using PolyPhen-2[296] and MutationTaster[297] suggested that the 4 Inuit specific missense variants are potentially deleterious (**Table IV**). Even with the exclusion of the Inuit high frequency variant *CPT1A* p.P479L (rs80356779), the Polyphen-2 scores of all the rare variants found in Nunavik Inuit and 1KGP Asian population were still significantly higher (Mann-Whitney U test, p=0.02) for Nunavik Inuit (**Table V** and **Figure S3**).

Mutation burden test and F-statistics of variants in the carnitine acyltransferase genes in Nunavik Inuit

In order to characterize the genetic profile of the carnitine acyltransferase genes in Nunavik Inuit, the mutation burden for each gene was calculated in the Inuit and compared to the mutation burden in 286 Asians (CHB and JPT) from 1KGP. Nunavik Inuit had significantly more mutated alleles per individual in the *CPT1A, CPT2* and *CRAT* genes than Asians (Binomial test, $p<0.01$). Permutation test showed similar results with marginally significant p values in *CPT2* and *CRAT* (not significant after *Bonferroni* correction) (**Table VI**).

F_{ST} and F_{IS} scores were calculated for 100 Nunavik Inuit and 83 HapMap Asians (CHB and JPT). The *CPT1A* p. P479L variant had the highest F_{ST} value (0.92), and the mean F_{ST} of the 4 Inuit specific variants was also statistically significant (0.84, $p<0.01$). However, the mean F_{IS} of all variants did not significantly deviate from zero after 1,000 randomizations (-0.0092) (**Table VII**).

Figure 10 depicted the frequencies and deleterious scores of the functional variants of the carnitine acyltransferase gene in different populations. The allele frequencies of all missense variants in European descendants (EVS database), Asians (1KGP) and the Nunavik Inuit were plotted against the PolyPhen-2 score of each variant. Variants with PolyPhen-2 score>0.8 (predicted to be highly deleterious) were presented at higher frequencies in the Nunavik Inuit compared to Asians and European descendants (**Figure 10**).

Common missense variants in the carnitine acyltransferase genes

HWE testing did not show a significant deviation for any of the exonic variants identified in the 5 genes. Haplotype analysis found that 3 common Nunavik Inuit *CPT1B* missense variants (rs470117, rs8142477 and rs3213445) were in an 8 kb LD block. The frequency of this haplotype in the Inuit (0.66) was significantly higher than in Asians (0.46) (Binomial test, p=7.501e-05). In *CPT2*, two common missense variants rs2229291 (p.F352C) and rs1799821 (p.V368I) are in a 1 kb LD block (**Figure 11**). These variants along with the Nunavik Inuit specific variant p. R477W form four major haplotypes in the studied cohort: T-G-C (45%), T-A-C (24%), G-A-C (28%) and T-G-T (3%), respectively. While in Europeans and Africans, only the T-G-C and T-A-C haplotypes were found. In Nunavik Inuit the G-A-C haplotype frequency is 28% compared to 10% in Asians[298], and absent in Europeans (Binomial test, p=3.481e-07).

Discussion:

Carnitine acyltransferase gene variants in the Nunavik Inuit

Our results indicate that Nunavik Inuit are genetically distinct from European, African and Asian populations, with the closest relationship to Asians. In 100 Nunavik Inuit we found three novel missense variants and a known Inuit specific missense variant in 4 carnitine acyltransferase genes. The rare missense mutations are significantly more deleterious and their frequencies are significantly higher in Nunavik Inuit, compared to other populations. Interestingly, we didn't find *CROT* missense variants in either Nunavik Inuit or Asians.

Mutation burden testing shows that mutations in the CPT1A, CPT2 and CRAT genes are significantly more frequent in Nunavik Inuit compared to Asians. The mean F_{ST} value (0.18) of all variants indicates a great genetic differentiation between Nunavik Inuit and Asians ($F_{ST}>0.15$); it is also noteworthy that the 4 Inuit specific variants are indicating an even higher degree of differentiation for this population ($F_{ST}>0.25$) (**Table VII**). However the F_{IS} value shows no statistical significance, which may due to the limited number of variants available. Nevertheless, the negative mean F_{IS} value suggests that there could be an excess of heterozygotes within the populations.

Nunavik Inuit population specific variants

CRAT and *CPT2*

The *CRAT* p.S78F mutation is located at the splicing site and is predicted to affect splicing, potentially to leading to a loss of function of CrAT and carnitine acetyltransferase deficiency.

Moreover, in a muscle-specific CrAT knockout mouse model, CrAT acts as a modulator of whole-body glucose homeostasis and metabolic flexibility[299]. The Inuit appear to tolerate drastic changes in metabolic homeostasis, hence the loss of CrAT function may be beneficial to them. Symptomatic CPT-II deficiency is usually caused by homozygous or compound heterozygous mutations in *CPT2*. The *CPT2* p.R477W mutation found in the Nunavik Inuit is likely to lead to the loss of function as it is located in a highly conserved region where mutations known to cause CPT-II deficiency occur[300; 301]. It is possible that heterozygote loss of function mutations in *CPT2* may have a beneficial effect on the enzyme activity in Nunavik Inuit.

The high frequency *CPT1A* p.P479L variant

In general, missense mutations are rare in *CPT1A*; yet in some Inuit populations, the p.P479L loss of function mutation has a high frequency, ranging from 44% to 83% from Alaska to Greenland[48; 226]. In our study of the Nunavik Inuit, the p.P479L mutation has an allele frequency of 95.5%, the highest reported to date. The frequency of this variant seems to increase from the west to the east along the arctic tree line and from inland to the shore (**Figure S4**), correlating with the migration timeline and with the higher consumption of animal fat of residents near the shore. The *CPT1A* p.P479L mutation is significantly more frequent among Nunavik Inuit compared to 243 Kivalliq Inuit (western Nunavut) ($\chi 2=15.629$, $p=0.0004$)[48]. Interestingly, the *CPT1A* p.P479L mutation is absent from all other worldwide populations, including East Asians (**Table IV**). This variant was initially believed to cause CPT-I deficiency, since the first discovery was made in a Canadian First Nation man with myopathy[52], a typical subtype of CPT1A deficiency. Therefore, this variant has been included in the Alaska newborn screening protocol[302]. In addition, in Canadian Inuit and First Nation families with severe CPT-I deficiency, the only missense mutation found in the *CPT1A*

gene was the homozygous p.P479L variant[53]. It was further demonstrated that the presence of the p.P479L variant in both *CPT1A* alleles resulted in reduced CPT-I activity in cultured fibroblasts and affected malonyl-CoA interaction with CPT1A[303]. However, other studies of this variant in Yup'ik Eskimos and Greenland Inuit yielded different results. In these studies, the L479 allele was reported to be associated with infant mortality[51], with impaired fasting tolerance[50], reduced adiposity[304] and with higher levels of HDL-cholesterol and apoA-I cardioprotective factors[305]. Nevertheless, it is important to note that Yup'ik Eskimos and Canadian Inuit are genetically different[26]. The reduced CPT1A activity associated with the homozygous state of p.P479L variant in Inuit may indicate two scenarios: A) lower activity of the p.P479L variant is beneficial for a state of permanent ketoadaptation in Inuit[53] or B) another Inuit specific variant in the regulatory region may serve as a rescue factor to the enzyme activity.

Although different disorders and traits associated with this variant were reported in different studies, the presence of this variant in very high frequency in Nunavik Inuit suggests that it is likely to be beneficial in this population. For example, it may help the arctic residents to adapt to the extremely cold environment and/or their ketogenic diet. This assumption is supported by the mean age difference found in previous studies and in the current study. Previously, it was reported that the p.P479L variant frequency is higher among Inuit newborns and children[48; 51]. However, the current study shows that among older populations (mean age of 52 years at enrollment) the L479 allele remains predominant. Furthermore, infant mortality rates are higher in Nunavik than in Nunavut (25 vs 14.6 per 1,000 live births) and Quebec (5 per 1,000 live births), but lower in Kivalliq (32.3 per 1,000 live births)[306]. These conflicting data suggest that further investigation is needed to determine whether there is an association between the p.P479L variant and infant death. Since the p.P479L variant is absent from Chinese and Japanese populations, it probably occurred

more recently, after the migration across the Bering Strait of their Asian ancestors. Genotyping the p.P479L variant in Siberian and Mongolian populations will be interesting to trace the origin and the occurrence of this variant.

CPT1B and *CPT2* haplotypes in the Nunavik Inuit population

The mutation burden and tolerance tests did not show any excess of rare missense *CPT1B* variant in Nunavik Inuit. Nevertheless, the relative frequencies of different *CPT1B* locus rs470117/rs8142477/rs3213445 containing haplotypes found in the Nunavik Inuit are different from those seen in Asian populations. Interestingly, this haplotype is in the same LD block as the SNP rs5770917, which was reported to be associated with narcolepsy in the Japanese population[307; 308]. Given that the Inuit live in the far north with the midnight sun and the polar night, variants in this locus may therefore have some roles in sleep, possibly benefiting the Inuit while predisposing to narcolepsy in Japan.

CPT2 missense mutations are rare in the general population, suggesting that the variations in the *CPT2* gene seen in Nunavik Inuit may be functional, possibly related to the energy metabolism requirements unique to this population. The haplotype containing p.F352C and p.V386I variants was previously named as a thermolabile CPT-II variant[309] with decreased CPT-II activity. It was further reported as a risk factor for infection-induced acute encephalopathy and for continuous high-grade childhood fever, which leads to a systemic and metabolic energy crisis in Japanese and Chinese populations[298; 309; 310]. The *CPT2* p.F352C variant was also reported in three individuals from one Inuit family with CPT-II deficiency[53]. However, there was no clinical description of acute encephalopathy in this family. Since this variant is temperature sensitive, it may act differently or even lead to higher enzyme activities in cold temperatures, which may explain their increase of the thermolabile haplotype frequency in the Inuit. Of note, encephalopathy

was not reported in the Nunavik Inuit cohort, suggesting that the same genetic variation may be associated with different phenotypes in different populations.

In this study, we observed an increased frequency of rare divergent variants in the carnitine acyltransferases family of genes in Nunavik Inuit in Quebec, as compared to other populations. There are few missense mutations in these genes, suggesting that they do not tolerate variation. As Asians are thought to have common ancestors with the Inuit[1], the excess of variants in these genes in Nunavik Inuit also suggests an effect of selective pressure. It is possible that these variants are related to their high fat diet, as the carnitine acyltransferase genes are essential for fat metabolism.

Since data on metabolic measurements was not available in our cohort, we could not determine whether the high frequency of variations in the carnitine acyltransferase genes in Nunavik Inuit affects their enzyme levels and activities. Our cohort was comprised of healthy Nunavik Inuit individuals and individuals with a family history of brain aneurysms; none of them showed any of the severe symptoms caused by CPTs or CrAT deficiencies. We hypothesize that these variants, while being part of the Inuit 'healthy genomes', could be harmful in other populations. Further studies in other Inuit populations, with the inclusion of measurements of these enzymes' levels and activity are necessary to confirm our results and conclusions.

Acknowledgements: We would like to thank the Northern Quebec Module (MNQ), which operates under the authority of the Nunavik Regional Board of Health and Social Services, for their support and their recruitment of Nunavik Inuit samples of this study. We also wish to thank Dr. Damian Labuda (CHU Sainte- Justine, Université de Montréal), Dr. Simon Girard (McGill University and Genome Quebec) and Dr. Ziv Gan-Or (Montreal Neurological Institute) for their valuable suggestions towards this work.

Figures

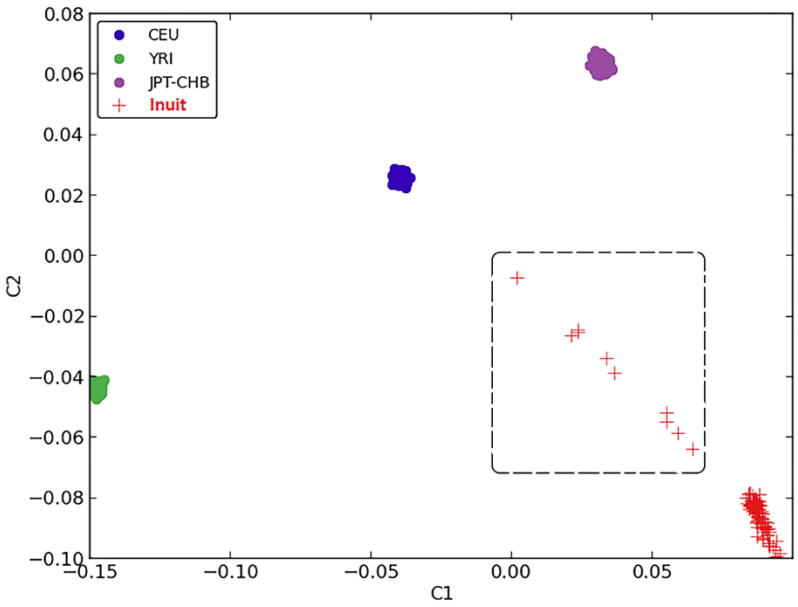

Figure 9. MDS plot showing the distinctive ethnicity relationships of CEU, YRI, JPT-CHB and Nunavik Inuit.

Individuals in the dashed-line box have a mixture of both Inuit and European genomes.

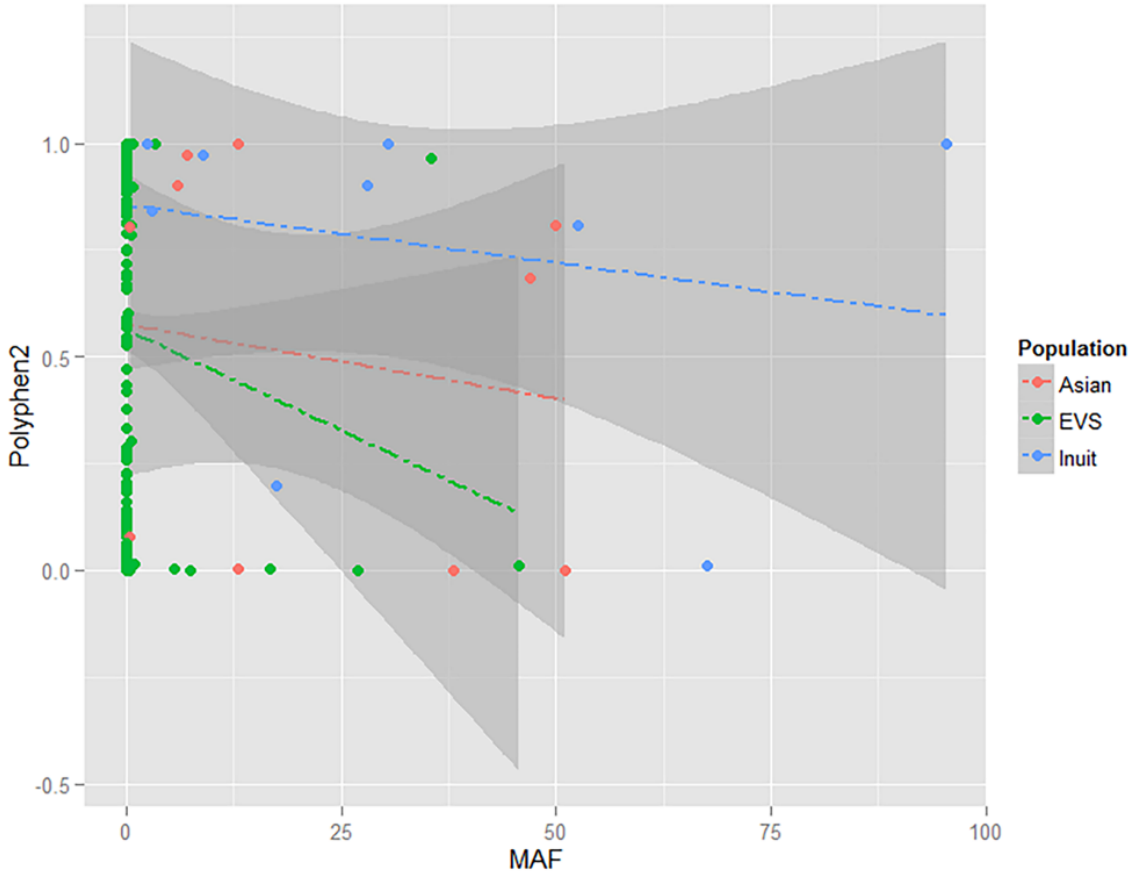

Figure 10. Scatterplot of the frequencies and deleterious scores of *CPT1A*, *CPT1B*, *CPT1C*, *CPT2* and *CRAT* missense variants.

Variants were extracted from 286 Asians from 1KGP, EVS populations comprising 4,294 European descendants and 100 Nunavik Inuit.

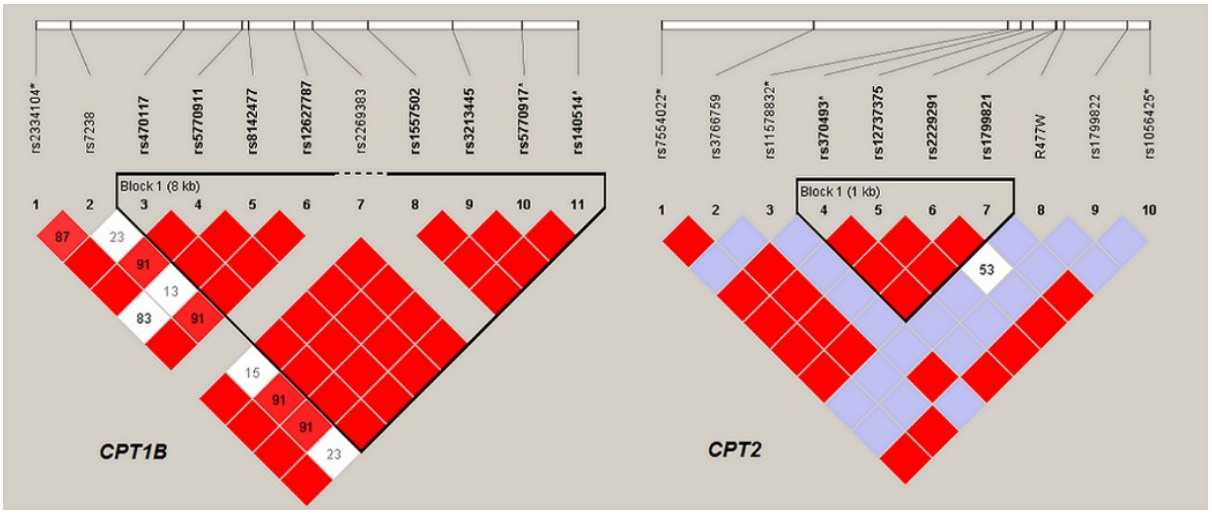

Figure 11. Pairwise linkage disequilibrium (LD) diagram for *CPT1B* and *CPT2* in Inuit.

LD block was delineated using confidence intervals (Gabriel et al), and 0.6–0.98 for strong LD. Variants of MAF<0.05 were indicated with dashed-line. Variants were extracted from exome sequencing and SNP array data, with the asterisk indicated variants only genotyped by the SNP array. D value <1 was shown in the box, with red color indicated LOD ≥2.

Tables

Table II. Genomic features of carnitine acyltransferase genes.

	Encode	**Genomic region of longest isoform (number of isoforms)**	**Number of exons**	**HGMD** mutation numbers from of all isoforms**	**Loss of function variant* from EVS*****
CPT1A	carnitine palmitoyltransferase -I (CPT-I)liver-type	chr11:68522088-68609399 (2)	20	29 (missense) 1 (splicing) 5 (indel)	3
CPT1B	CPT-I muscle-type	chr22:51007290-51017096 (4)	21	2 (missense) 1 (splicing)	9
CPT1C	CPT-I brain-type	chr19:50194365-50216988 (2)	20	N/A	4
CPT2	carnitine palmitoyltransferase –II (CPT-II)	chr1:53662101-53679869 (1)	5	65 (missense) 4 (splicing) 20 (indel)	3
CRAT	carnitine acetyltransferase (CrAT)	chr9:131857073-131873070 (1)	15	N/A	3
CROT	carnitine octanoyltransferase (CrOT)	chr7:86974951-86989425 (3)	18	N/A	12

*****Including nonsense mutations, frameshift mutations, splicing donor and acceptor site mutations.

******Human Gene Mutation Database, including all published gene mutations responsible to human inherited disease.

*******NHLBI Exome sequencing project exome variant server, including approximately 4,294 European Americans and 2,200 African Americans.

Table III. Coding variants of carnitine acyltransferase genes discovered in Nunavik Inuit.

Gene	Variant	Variation type	Annotation
CPT1A	p.V616V	synonymous	novel
CPT1A	p.P479L	missense	rs80356779
CPT1A	p.F417F	synonymous	rs2228502
CPT1B	p.E531K	missense	rs470117
CPT1B	p.S427C	missense	rs8142477
CPT1B	p.I66V	missense	rs3213445
CPT1C	p.T265M	missense	novel
CPT2	p.F352C	missense	rs2229291
CPT2	p.V368I	missense	rs1799821
CPT2	p.R477W	missense	novel
CRAT	p.A603P	missense	rs17459086
CRAT	p.A575A	synonymous	rs375414636
CRAT	p.S78F	missense_splicing	novel

Table IV. Variant frequencies and deleterious score predictions of carnitine acyltransferase genes.

Gene	Protein coding variants	SNP	MT	PP2	Frequency* in Nunavik Inuit (100)	Frequency in HCB-JPT (286)**	Frequency in CEU (178)**	Frequency in YRI (250)**
CPT1A	p.P479L	rs80356779	0.997899	1	95.5%	0	0	0
CPT1B	p.E531K	rs470117	0.606166	0.303	30.5%	48.3%	46.5%	8.8%
CPT1B	p.S427C	rs8142477	0	0	67.5%	51.7%	93.3%	25.8%
CPT1B	p.I66V	rs3213445	0.251908	0	17.5%	35.8%	4.9%	10.2%
CPT1C	p.T265M		0.999717	0.998	0.5%	0	0	0
CPT2	p.F352C	rs2229291	0.999596	0.999	28.0%	20.0%	0	1.3%
CPT2	p.V368I	rs1799821	0.06145	0.001	52.5%	77.5%	54.9%	25.2%
CPT2	p.R477W		0.997459	1	3.0%	0	0	0
CRAT	p.A603P	rs17459086	0.993409	0.013	9.0%	7.0%	1.7%	0.8%
CRAT	p.S78F		0.999657	0.906	2.5%	0	0	0

*Based on the derived allele from the reference genome.

**From 1000 genome database.

Table V. Deleterious scores of all rare missense variants in carnitine acyltransferase genes found in Nunavik Inuit and 1KGP Asians.

	PolyPhen v2	**Derived allele frequency (%)**
Nunavik Inuit (100)		
***CPT1A:* p.P479L**	1	95.5
***CPT1C*: p.T265M**	0.998	0.5
***CPT2:* p.R477W**	1	3
***CRAT*: p.S78F**	0.906	2.5
1KGP Asians (286)		
***CPT1A:* p.I491T**	0	0.35
***CPT1B*: p.C659W**	1	0.35
***CPT1C:* p.R514Q**	1	0.35
***CPT1C:* p.Q97H**	0.08	0.35
***CPT2:* p.S122F**	0.018	0.6
***CRAT*: p.V411M**	0.803	0.35

Table VI. Mutation burden scores of identified rare mutations (MAF<0.01) of carnitine acyltransferase genes in Nunavik Inuit and 1KGP Asians.

Gene	Variants	Av # variants alleles/individual		Inuit/Asian ration	Base-pair sequenced	P*	P**
		Inuit (100)	Asians (286)				
CPT1A	*p.P479L****, p.I491T	1.91	0.007	272.80	2319	**2.2e-16**	**0.000006**
CPT1B	p.C659W	0	0.007	0.00	2316	1	0.75
CPT1C	*p.T265M****, p.Q97H, p.R514Q	0.01	0.014	0.71	2409	1	1
CPT2	*p.R477W****, p.S122F	0.06	0.017	3.53	1974	**0.003**	0.03689
CRAT	*p.S78F****, p.V411M	0.05	0.007	7.14	1815	**0.00076**	0.01305

*Binomial test, Bonferroni corrected, significant p<0.01

**Empirical p-value of permutation test, 105 permutations

***Variants only identified in Inuit

Table VII. The F_{ST} and F_{IS} value of 13 variants in the population containing Nunavik Inuit and HapMap Asians.

Variant	SNP	F_{ST}	F_{IS}
***CPT1A*: p.V616V**		0.0196	-0.04
***CPT1A*: p.P479L**	rs80356779	**0.9171***	-0.0452
***CPT1A*: p.F417F**	rs2228502	0.0506	-0.1067
***CPT1B*: p.E531K**	rs470117	0.0414	-0.0273
***CPT1B*: p.S427C**	rs8142477	0.034	0.0382
***CPT1B*: p.I66V**	rs3213445	0.0453	0.0114
***CPT2*: p.F352C**	rs2229291	0.0298	-0.0925
***CPT2*: p.V368I**	rs1799821	0.0546	0.0349
***CPT2*: p.R477W**		0.0146	-0.0297
***CRAT*: p.A603P**	rs17459086	0.0045	-0.0227
***CRAT*: p.A575A**	rs375414636	0.0097	-0.0196
***CRAT*: p.S78F**		0.0122	-0.0246
***CPT1C*: p.T265M**		0.0024	-0.0048
Mean (all variants)		0.1833	-0.0092
Mean (missense variants)		0.1907	-0.0053
Mean (Inuit specific variants)		**0.8401***	-0.0393

F_{ST}: The fixation index, which is the proportion of total genetic variance contained in a subpopulation relative to the total genetic variance. The value of F_{ST} ranges from 0 to 1, the higher value implies higher degrees of population differentiation.

F_{IS}: The inbreeding coefficient, which is the proportion of genetic variance in the subpopulation contained in an individual.

*Statistical significance, 1000 randomizations

CHAPTER 3: UNCOVERING MISSING HERITABILITY OF INTRACRANIAL ANEURYSM IN THE FRENCH-CANADIAN FOUNDER POPULATION

3.1: PREFACE

IA is a common form of the cerebrovascular disorder and a number of populations are known to present at higher risks to develop this condition[89; 211]. In this chapter, we furthered our efforts to identify IA susceptible genes. Here we focused on the French-Canadians, an important founder population of Quebec. Similar to what has been observed for Nunavik Inuit, French-Canadians also present an increased prevalence of IA, especially for the familial form.

Because of the heterogeneity and complexity of IA, only two studies[218; 219] have thus far used a WES approach; however, their results failed to explain most IA cases. In Chapter 3.2, we describe a study that successfully used WES to identify an IA genetic risk factor, *RNF213*; the study relied not only on WES but also on targeted resequencing and a protein assay. Variants of *RNF213* were previously established to cause another cerebrovascular condition known as moyamoya disease (MMD). In this study, the French-Canadian IA patients from large families were chosen as the discovery cohort, and *RNF213* was identified to bear significant increased genetic variation burden among all genes from WES. Targeted resequencing of *RNF213* was subsequently performed in a larger cohort of IA cases (70% of which were familial) and additional rare missense mutations were identified, in turn representing a higher genetic burden compared to controls. ATPase assays were performed to measure two of the IA variants located in the functional domain of the protein expressing ATPase activity. The measurements revealed these two IA

variants to produce a gain of ATPase activity; and interestingly this was the opposite of what the MMD variants produce. Distinct *RNF213* variants can therefore affect cerebrovascular development in a manner that will lead to distinct cerebrovascular disorders.

On the other hand, genome-wide association studies (GWAS) are still considered to represent the most powerful approach for the identification of IA risk genes. In Chapter 3.3, we present a GWAS study focusing on familial IA patients of French-Canadian origin. Unlike previous large GWASs, we focused on the French-Canadian founder population to increase the detection power. With our cohort aggregated with familial cases, it is likely to discover new candidate loci harboring risk genes for IA that could further explain the missing heritability of this cerebrovascular disorder. In this study, 257 French-Canadian IA patients and 1,992 controls were genotyped using custom designed Illumnia NeuroX SNP-chip of approximately 710,000 variants backbone and 24,000 variants focusing on neurological diseases. After imputation, we discovered a locus with genome-wide significance at 3p14.2. The SNP within this top association was located in an intronic region of *FHIT* (rs1554600, $p=4.66\times10^{-9}$), previously suggested to be associated with the risk of hypertension in Quebec, Canada. Another interesting gene *CCDC80* (rs78125721, $p= 4.77\times10^{-7}$) was located in another promising locus of 3q13.2; WES of 32 French-Canadian IA cases and 106 control individuals revealed a significantly increased exonic variant burden in *CCDC80* across IA cases. The product of *CCDC80* is noticeably expressed in vascular smooth muscle cells and is involved in cell adhesion during differentiation of fibroblast. Both genes may contribute additionally to the vascular development and risk for IA in French-Canadians.

3.2: RNF213 IS ASSOCIATED WITH INTRACRANIAL ANEURYSMS IN THE FRENCH-CANADIAN POPULATION

Manuscript: *RNF213* is associated with Intracranial Aneurysms in the French-Canadian population

American Journal of Human Genetics, 2016 Nov 3; 99(5):1072-1085. doi: 10.1016/j.ajhg.2016.09.001

Aurthors and affiliations:

Sirui Zhou[1,2], Amirthagowri Ambalavanan[1,3], Daniel Rochefort[1], Pingxing Xie[1,4], Cynthia V. Bourassa[1], Pascale Hince[1], Alexandre Dionne-Laporte[1], Dan Spiegelman[1], Ziv Gan-Or[1,3,8], Cathy Mirarchi[1], Vessela Zaharieva[1], Nicolas Dupré[5], Hatasu Kobayashi[6], Toshiaki Hitomi[6], Kouji Harada[6], Akio Koizumi[6], Lan Xiong[2,7], Patrick A. Dion[1,8], Guy A. Rouleau[1,8,*]

[1]Montreal Neurological Institute and Hospital, McGill University, Montréal, QC, Canada, H3A 2B4.

[2]Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal, QC, Canada, H3T 1J4.

[3]Department of Human Genetics, McGill University, Montréal, QC, Canada, H3A 0G4.

[4]Faculty of Medicine, McGill University, Montréal, QC, Canada, H3A 0G4.

[5]Department of Neurological Sciences, CHU de Québec and Department of Medicine, Faculty of medicine, Laval University, Quebec, QC, Canada, G1V 0A6.

[6]Department of Health and Environmental Sciences, Graduate School of Medicine, Kyoto University, Yoshida Konoecho, Kyoto, Japan, 606-8501.

[7]Centre de recherche, Institut universitaire en santé mentale de Montréal, QC, Canada, H1N 3M5.

[8]Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada, H3A 0G4.

Authors contribution:

Study design: SZ, GAR, PAD, LX

Wet-lab experiment: SZ, DR, PH

Bioinformatic analysis: SZ, AA, PX, ADL, DS

Sample recruitment: CVB, ZGO, CM, VZ, ND

Vector construct: HK, TH, KH, AK

Manuscript writing: SZ, LX, PAD, GAR

ABSTRACT

Intracranial aneurysms (IAs) are the result of focal weakness in the artery wall and have a complex genetic makeup. To date, genome-wide association and sequencing studies have had limited success in identifying IA risk factors. Distinct populations, such as the French-Canadian (FC) population, have increased IA prevalence. In our study, we used exome sequencing to prioritize risk variants in a discovery cohort of six FC families affected by IA, and the analysis revealed an increased variation burden for *ring finger protein 213* (*RNF213*). We re-sequenced *RNF213* in a larger FC validation cohort, and association tests on further identified variants supported our findings (SKAT-O, p=0.006). RNF213 belongs to the AAA+ protein family, and two variants (p.Arg2438Cys and p.Ala2826Thr) unique to affected FC individuals were found to have increased ATPase activity, which could lead to increased risk of IA by elevating angiogenic activities. Common SNPs in *RNF213* were also extracted from the NeuroX SNP-chip genotype data, comprising of 257 FC IA-affected and 1,988 control individuals. We discovered that the non-ancestral allele of rs6565666 was significantly associated with the affected individuals (p=0.03), and it appeared as though the frequency of the risk allele had changed through genetic drift. Although *RNF213* is a risk factor for moyamoya disease in East-Asians, we demonstrated that it might also be a risk for IA in the FC population. It therefore appears that the function of RNF213 can be differently altered to predispose distinct populations to dissimilar neurovascular conditions, highlighting the importance of a population's background in genetic studies of heterogeneous disease.

INTRODUCTION

Intracranial Aneurysms (IA) (MIM105800) are vascular abnormalities characterized by dilations or ballooning of intracranial arteries. The pathology of IA is not clear, however the lesions are usually characterized by very thin (or absent) tunica media and internal elastic lamina within the arterial walls at the IA sites[311]. The worldwide prevalence of IA is 1-3%[102; 312]. Moreover this condition entails severe consequences, particularly when subarachnoid hemorrhages (SAH) occur (incidence rate 0.5-1%[98]) as 30-45% of SAH are fatal within 30-days[313]. The annual incidence of SAH is 4-9 per 100,000 worldwide and 80% of all spontaneous SAH are caused by the rupture of aneurysm[314]. The formation and rupture of IA are related to complex risk factors, including smoking and alcohol consumption, hypertention[315] and other vascular diseases[316], a very small percentage of them were related with infections[317] and trauma[318]. Many studies provided evidence suggesting a large fractions of IA and SAH cases involve underlying genetic risk factors. First-degree relatives of individuals with aneurysmal SAH have a 4 to 7 fold increased risk of being affected by comparison to the general population[319]. There have been several genome wide association (GWA) studies examining sporadic IA and these identified several candidate loci (e.g. 2q33.1, 8q11.23, 9p23.1[207; 214]; 18q11.2, 13q13.1, 10q24.32[320]; 7p21.1[215]; regions near rs1930095, rs7781293, rs7550260, rs9864101[208]; and 4q31.22[211]). Unfortunately, follow-up replication studies suggests those loci are unlikely to explain a large fraction of IA cases. It has been reported that the French-Canadian (FC) population has a higher IA and SAH incidence rates and affected individuals usually aggregated in families, especially large pedigrees[321]. In a study conducted in the Saguenay-Lac-Saint-Jean region of Quebec, Canada, it was observed that 30% of IA affected individuals had a family history[322]. It is believed that IA is genetically heterogeneous and therefore population specific variants might play a big part of the pathogenesis, especially in familial

aggregates. In order to explore those hypotheses, we conducted whole exome sequencing (WES) and target sequencing experiment in a group of familial IA subjects of FC descent. We focused our attention to the identification of rare, or FC specific, variations as genetic factors that might explain the high prevalence and familial aggregation of IA in this population.

MATERIALS AND METHODS

Discovery cohort

The discovery cohort of this study included 26 French-Canadians (from 6 families) who had received a positive IA diagnosis (**Figure 12**); diagnostics were confirmed either by magnetic resonance angiography (MRA) (1.5T), or by surgical confirmation (clipped or coiled). Subjects were excluded if they had received a personal diagnosis, or had a history of polycystic kidney disease (MIM173900), Ehlers-Danlos syndrome type IV (MIM130050), neurofibromatosis type 1 (MIM162200), Marfan syndrome (MIM154700) or other intracranial vascular malformations including moyamoya disease (MIM607151).

Written consent was obtained for all the participants and the study was approved by Comité d'éthique de la recherche du Centre hospitalier de l'Université de Montréal (ND 04.101, Québec, Canada).

Population control cohort

Exome population controls for this study were selected from our in-house exome database, which includes 189 healthy FC individuals. Additionally, the SNP-chip genotype data for 1,988 unrelated FC control individuals with no obvious cerebrovascular disease were also included for this study.

Whole exome sequencing (WES) and variants prioritization

Genomic DNA for each individual was extracted from peripheral blood lymphocytes using Gentra Systems PUREGENE DNA purification kit (Qiagen). A 50 µl DNA sample at a concentration of 100 ng/µl from each sample was captured by Agilent SureSelect V4 capture kit. The library was subsequently sequenced at 100 bp pair-end using Illumina Hiseq 2000, with 3 samples per lane to ensure an average coverage depth of 100-fold. All high-throughput and Sanger sequencings were performed at the McGill University and Génome Québec Innovation Centre.

Raw fastq files were aligned to NCBI human reference GRCh37 using Burrows-Wheeler Aligner (BWA)[323], with all PCR duplicates removed from the alignments. The aligned reads were converted to binary format for further analysis using Sequence Alignment/Map (SAM)[288] tools. Single nucleotide variant (SNV) and insertion/deletion (indel) calling done by UnifiedGenotyper from Genome Analysis Toolkit (GATK)[267] version 2.7. Variant annotation was performed using ANNOVAR[289] program with references to GRCh37/hg19, dbSNP version 132, 1000 Genomes project (1KGP) (2012 data release)[290], 69 Complete Genomics (2012 update) and exome variant server (EVS) with approximately 6,500 exomes (NHLBI-ESP project, 2013 update). Finally, variant segregation analysis was performed using an in-house segregation program, using more than 1,000 exomes of different ethnicities from our lab as controls.

The sequencing quality were determined by GATK's DepthOfCoverage Walker. QC step excluded variants with sequencing depth <10 or with genotype quality <90, and of those presented in a selected of frequently mutated genes, pseudo-genes and genes in repetitive regions. After QC, preliminary analysis focused on variants that potentially affect protein function (missense, splicing, small indels). Variants from all aforementioned databases were removed, to focus on French-Canadian specific variants. Polyphen-2[296] and GERP++[324] were used to predict the deleterious level of the functional variants.

Subsequently, French-Canadian specific, potential deleterious variants shared by two or more affected individuals from each family were prioritized for further analysis, and genes harboring two or more of those variants are further selected of interest.

Variants from 106 French-Canadian exome controls from those genes of interest were selected by the same criteria, and gene-based burden test was performed to those genes using variable thresholds method[325] implemented in Variant Association Tools (VAT)[326].

Validation cohort and target resequencing

After the prioritization of variants and the selection of the most promising candidate gene, the whole coding region of two isoforms of the *RNF213* gene (MIM613768) (NM_020954 and NM_001256071) were sequenced in a validation cohort of 223 affected individuals and 88 IA negative controls. 74 of the subjects were sporadic and rest were with family history. The mean age of onset for familial IA were 53.5±12 years compared to sporadic IA which were 55.4±9 years. Out of the individuals with detailed clinical information, 37.4% of subjects with family history were indicated to have multiple aneurysms, compared to 29% in sporadic cases. On the other hand, 40.8% of the familial IA subjects were indicated to have hypertension, while among sporadic subjects 47.7% were shown with high blood pressure. Other environmental factors also seemed to play a larger part in sporadic cases, such as drinking (22.4% familial vs 40% sporadic) and smoking (60.5% familial vs 76.9% sporadic) (**Table S1**). The targeted resequencing was done using Fluidigm Access Array™ System. 122 pairs of primers were designed to cover a total of 58 exons of *RNF213*, with each fragment under 300 bp in length (**Table S2**). 50ng of DNA for each sample were used for performing multiplex PCR and products were sequenced paired-end on Illumina MiSeq platform.

Bioinformatic analysis of the Fluidigm sequencing data was done to first remove bad QC reads and adaptor sequences. Subsequently the reads were aligned against GRCh37 using the Burrows-Wheeler Aligner (BWA) v0.7.5[327] and alignments of each sample were combined. UnifiedGenotyper[267] implemented in the Genome Analysis Toolkit (GATK) were used to perform variant calling in the coding intervals of *RNF213*. The annotation and segregation steps were made in a manner that was similar to what was done for WES. All variants were validated by IGV[328] and rare (MAF $\leq$ 0.01) missense variants were further validated by Sanger sequencing.

Genotyping RNF213 common SNPs

SNP genotyping data from NeuroX SNP-chip (Illumina Human OmniExpress bead chip as backbone) was obtained for 257 IA subjects and 1,988 controls, which contains 719,885 SNPs. Principle component analysis (PCA) was performed using smartPCA software (implemented in eigensoft)[329] to confirm the FC ethnicity of the samples used, and SNPs within the *RNF213* gene region was extracted from the genome wide data. Additionally, *RNF213* common SNPs from 33 different populations were also obtained from in-house database, Gene Expression Omnibus (GEO) database[16; 39; 330; 331] and 1000 genome (1KGP) phase 3 data[290] (**Table S3**).

Statistical and population analysis

In order to further select the genes that are most likely to be a risk candidate for IA, gene-based burden test was performed for genes prioritized in initial cohort using variable thresholds (VT) method[325] implemented in Variant Association Tools (VAT)[326].

Sequence Kernel Association Test (SKAT)[332] was performed to evaluate the effect of *RNF213* variants with weighted scores; this unified statistical test allows both rare and common variants to contribute to the overall statistic by calculating the logistic weights for each variant and apply them to the analysis using the following formula:

$$weights = \frac{e^{(par1-MAF)par2}}{1+e^{(par1-MAF)par2}}$$

SKAT optimal test (SKAT-O) was then performed to evaluate the statistical significance of *RNF213* functional variants between affected individuals and controls.

A Cochran-Armitage trend test was performed using Plink 1.9[273] to compare the allele distributions of all *RNF213* common SNPs in IA cases and controls; an adjustment was made for multiple testing. Pairwise F_{ST} were calculated using Arlequin 3.5[333] between different populations to measure the population differentiation according to two different sets of their *RNF213* variations: (1) *RNF213* common SNPs from FC IA subjects, controls and 33 other populations; (2) *RNF213* functional variants from FC IA subjects, controls and 25 populations from 1KGP phase III data. Variation burden of *RNF213* between FC and 1KGP phase III populations were also calculated by Variable Thresholds (VT) method[325] using Variant Association Tools (VAT)[326].

Expression vector design

The *RNF213* full-length cDNA cloned into pcDNA3.1+ was obtained from a previous study[334]. Two AAA+ module fragments (aa 2,370–2,632 and 2,717–3,004, respectively) were individually amplified using Gibson assembly adapted primers, with 1st AAA primer GGT TCC GCG TGG ATC CCC GGA ATT Cgt gcc ctt caa tgt cga ctt tga taa ac and GTC AGT CAC GAT GCG GCC GCT CGA Ggc gag tcc cgt ttt cat cta gg; and 2nd AAA+ primer GGT TCC GCG TGG ATC CCC GGA ATT Cag cag gct gct tct gga tg and GTC AGT CAC GAT GCG GCC GCT CGA Gtc tat ttg aag cct ttg ctg cag caa aga cc. The two AAA+ amplified fragments were assembled and cloned into EcoRI and XhoI sites of the pGEX-4T1 vector (GE Healthcare) using the Gibson assembly method[335]. All clones were sequence validated.

Mutagenesis

PCR site-directed mutagenesis was performed to introduce two variants p.Arg2438Cys and p.Ala2826Thr in the 1st AAA+ and 2nd AAA+ plasmids, respectively. Four additional mutations (p. Lys2426Ala, p. Glu2488Ala, p. Lys2775Ala and p. Glu2845Ala) described previously[336] which known to cause loss of function (LOF) of the AAA+ ATPase were also introduced as controls.

Protein purification

Both the wildtypes, variants and LOF controls of the two AAA+ fragments and empty pGEX-4T1 plasmid were transformed into *BL21-CodonPlus* cells, the protein was inducted at 16°C for overnight and N-terminal glutathione S-transferase (GST)-tagged protein was extracted using Glutathione Sepharose 4B beads (GE Healthcare), in PBS buffer contain PMSF protease inhibitor. PBS was carefully removed after the protein extraction, GST protein was eluded using buffer contains 50 mM Tris-HCl and 10 mM reduced glutathione. Bradford protein assay was used for protein quantification.

ATPase assay

ATPase assay was performed using BIOMOL Green (Enzo life sciences) for the detection of free phosphate. The GST-tagged wildtypes, variants and LOF controls were incubated in a 50 µl buffer containing 300 mM KCl, 10 mM $MgCl_2$, 50mM HEPES-KOH at PH 7.5 and 5 mM ATP at 37°C for 20-40 min. The reaction was terminated by adding 100ul BIOMOL Green reagent, and incubated in room temperature for 10 min. OD_{630nm} was measured for the wildtypes, the variants, and the LOF controls and standard curve was made using phosphate standard based on the manufacturer's protocol. ATPase activity (Vmax) was measured by PO_4 quantity (nmol) per mg of protein per minute of reaction.

RESULTS

Variant Segregation analysis and RNF213 recurrent mutations

The average base depth of coverage of the 26 IA samples from the discovery cohort was 101-fold and 89.5% of the total target region was covered at 20-fold (**Table S4**). After QC and variant prioritization, 79, 109, 103, 15, 99 and 67 FC specific deleterious variants segregated across more than two affected individuals of each of the six families (**Table VIII, row G**). 35 out of these variants defined 19 genes which presented variants in more than one family, with 11 out of these 19 genes contain 25 distinct variants (**Table IX**). Interestingly, one of the 11 genes, *RNF213*, revealed 5 distinct variants in 4 families (**Table IX**, **Figure 12**). These *RNF213* (NM_001256071.2) variants were a nonsense mutation c.11413del; p.Glu3806Argfs*27, an exonic splicing mutation c.2017C>T; p.Arg673Trp and 3 missense mutations c.13577T>C; p.Ile4526Thr; c.6980A>G; p.Asn2327Ser) and c.3134C>T (NM_020954.3); p.Ser1045Leu (NP_066005.2). Only the c.6980A>G and c.3134C>T variants were observed once each in 189 FC descent control individuals. Variation burden of the 11 genes were displayed in **Table X**. After Bonferroni correction for multiple testing, only *RNF213* still differ significantly between subjects and controls (p=0.013).

Target re-sequencing validation

After Fluidigm capture and sequencing, the average coverage for the *RNF213* targeted region of each individual was calculated to range between 6,828 and 227,160-fold. Only 4% of individuals had ≤ 70% of the total target region covered at less than 100-fold.

272 exonic variants were first observed across the validation cohort of 311 samples (223 cases and 88 controls) and after IGV examination, 142 variants were used for the study (the remaining variants were deemed to be false positives because they were either located in adaptors or PCR duplicates). Non-coding and synonymous variants were further removed from the 142

variants and 60 variants were left for the final analysis. Additionally, 44 functional *RNF213* variants were also found in 189 French-Canadian population controls.

SKAT analysis and RNF213 rare variants in FC

In total, 72 functional (missense, indel, splicing) *RNF213* variants were observed in FC cases and controls, which including 233 IA subjects (10 selected from nuclear families of the discovery cohort and 223 affected individuals from the validation cohort) and 277 controls (88 from validation cohort and 189 from exome population controls) (**Table S5**). Hardy–Weinberg equilibrium (HWE) and missingness test was performed using Plink 1.9, no variant was removed.

The comparison of *RNF213* functional variants using a SKAT-O test with logistic weights showed significant different allelic distributions between 233 FC cases and 277 controls (P=0.008); the difference was even more significant when only rare functional variants (p=0.006) were compared. When the comparison of FC cases was examined using only 88 IA negative controls, the difference remained significant (p=0.018); if sex was added as covariate a slight increase of difference emerged (p=0.013).

It is also noteworthy that among the 34 rare functional variants of *RNF213* that were observed in 233 IA cases, 18 were absent from all controls (**Table S5**). 13 of which were predicted to be probably deleterious (Polyphen2 score > 0.8 or GERP++>2) (**Table XI**). On the other hand, 30 rare functional variants were observed in 277 controls, 14 were absent from affected individuals, from which only 4 were predicted deleterious (**Table S5** and **Table XI**). Hence there is a significant difference between the number of deleterious variants observed in only cases and controls (p=0.00006, binomial test). Twenty-five FC IA cases from 249 discovery and validation cohorts carry 14 rare deleterious *RNF213* variants (13 plus p.Glu3806Argfs*27) that absent in FC controls, from which nine individuals had multiple IAs, eight had SAH, eight with hypertensions

and three cases with hypercholesterolemia from which two also had myocardial infarctions. Only 6 individuals had no other recoded clinical histories besides smoking.

RNF213 common variants in population-based study

Among 257 IA subjects and 1,988 controls with SNP-chip genotyping data, PCA analysis showed that 3 IA cases and 92 controls presented an admixture of East-Asian and African ancestry and these samples were excluded from subsequent analysis (**Figure S1**). We performed gene-wide association test on 38 polymorphic SNPs across the *RNF213* loci in 254 cases and 1,896 control individuals, and the SNP rs6565666 appears to significantly associated with IA cases (p=0.03, After Bonferroni Correction, Cochran-Armitage trend test) (**Table S6**).

The pairwise F_{ST} comparisons of 1KGP phase III populations with FC suggest the common SNPs of *RNF213* vary between populations. FC individuals have a higher number of distinct variations compared with South-Americans, followed by Africans and East-Asians. However, East-Asians are sharing less *RNF213* functional variations with the FC population than they are with the African and Caribbean populations (**Figure 13**). Overall 8 SNPs of *RNF213* appear to be outliers in the frequency changes across the worldwide populations and 4 of these SNPs were located in the C-terminal domain region, it also appeared that more functional variants of *RNF213* were possibly being affected by genetic drift (**Figure S2**).

A burden test comparing the FC population to other populations showed a similar trend as East-Asians appear to have the most significant difference of derived *RNF213* allele frequencies when compared with overall FC, followed by Peruvians, Indians, and Finnish. As expected, FC IA cases also have significantly increased burden of *RNF213* than control individuals (**Table XII**).

Functional characterization of selected RNF213 variants

Among the 14 RNF213 variants predicted to be deleterious and specific to IA cases only two variants, p.Arg2438Cys (c.7312C>T) and p.Ala2826Thr (c.8476G>A), were located in the two distinct AAA+ modules of the protein. These two variants were therefore the most suitable for a functional test measuring the ATPase activity of RNF213. As previously reported[336] this activity can be detected using a recombinant fragment containing amino acids from 2,370–2,632 and 2,717–3,004.

The baseline of ATP hydrolysis was determined by the GST protein. As a result, the average ATPase activity of the 1st AAA+ protein wildtype (A1WT) and the p.Arg2438Cys variant (A1M) were estimated to be 72.0 nmol/mg/min and 142.1 nmol/mg/min, respectively; while the LOF variant in walker A domain (A1WA) was estimated to be 12.2 nmol/mg/min, with the walker B LOF variant (A1WB) 23.1 nmol/mg/min. The average ATPase activity of the 2nd AAA+ protein wildtype (A2WT) and the p.Ala2826Thr variant (A2M) were estimated at 127.1 nmol/mg/min and 242.6 nmol/mg/min, respectively; while the LOF variants of walker A and B domains (A2WA and A2WB) were as low as 8.8 nmol/mg/min and 15.5 nmol/mg/min, respectively. There were significantly increases of ATPase activities for both variants compared to their respective wildtype proteins (Mann-Whitney-Wilcoxon Test, $p_{A1M\text{-}A1WT}$=0.015, $p_{A2M\text{-}A2WT}$=0.0003); the differences were also significant compared with the LOF variants (Mann-Whitney-Wilcoxon Test, $p_{A1M\text{-}A1WA}$=0.002, $p_{A1M\text{-}A1WB}$=0.002; $p_{A2M\text{-}A2WA}$=0.0007; $p_{A2M\text{-}A2WB}$=0.0009) (**Figure 14**).

DISCUSSION

A major weakness of using a GWAS approach in IA is that the effect of rare variants and population specific variants are not well addressed. Recently two studies using exome sequencing on multiple IA families attempted to address the issue. Though no recurrent mutations were found,

using different hypotheses and in different populations they respectively reported 68[219] and 78 genes[218]. Only one gene *ITGB6* (MIM147558) was found to be common across the two lists established by these two studies; nonetheless both studies did not deem this to be an interesting IA candidate gene. We believe genetic heterogeneity is likely at play in the development of IA in distinct populations. Therefore, we conducted this research exclusively in FC individuals and focused on finding genes that harbor founder mutations/haplotypes in FC as candidates for predisposition to IA. Although in the current study, exome sequencing results of the first six FC families strongly suggested *RNF213* deleterious rare variants predispose to IA in FC individuals, there are also some other interesting variants that worth noting - such as *ABCA10* (MIM612508) that is expressed predominately in intracranial vascular endothelia cells[337].

The functional characterization of RNF213 gene and variants

Mysterin/*RNF213* (ring finger protein 213) has been identified as an important candidate for moyamoya disease (MMD) in the Japanese population[334; 338]. MMD is a rare cerebrovascular disorder that is characterized by stenosis of the internal carotid arteries and abnormal angiogenesis. It is a population specific condition with an extremely low prevalence in Western populations. A founder mutation, c.14429G>A (rs112735431; p.Arg4810Lys, previously annotated as XM_005257545.3: c.14576G>A; XP_005257602.2: p.Arg4859Lys[338]) was found significantly and specifically associated with the disease in the East Asian population[334]. However, in a study looking at French MMD subjects there was no evidence linking *RNF213* to MMD[339], another recent study further reported *ZXDC* and *OBSCN* with higher variation burden in affected individuals of European descent, suggesting the complexities in the MMD genetic etiology[340].

RNF213 encodes an AAA type ATPase with E3 ubiquitin ligase activity and studies suggest a role in vascular wall construction[334]. A functional study focused on RNF213

p.Arg4810Lys suggested that the founder mutation is a MMD risk factor through reduced angiogenic activity[341; 342] and induced mitotic abnormalities[343]. When exploring the functions of RNF213, studies have found the gene to be linked with a variety of artery wall developments. Several studies were conducted on *Rnf213* KO mice, showing thinning of the intima and media layer after CCA ligation[344], thinning in vascular walls and increased *Mmp9* expression[345] or enhanced post-ischemic angiogenesis[346]; a recent report also showed that *RNF213* was associated with inflammatory responses and angiogenesis[342; 347]; all these results suggest that *RNF213* is involved in vascular remodeling processes. A recent study even showed that RNF213 is associated with aneurysm formation post anastomotic surgery[348].

In addition to MMD, *RNF213* has also been associated with other vascular disorders[349] such as fibromuscular dysplasia (MIM135580)[350], high blood pressure[351], intracranial major artery stenosis[352; 353], and heterogeneous intracerebral vasculopathy[354]. The p.Arg4810Lys RNF213 mutation is suggested to be a risk factor for MMD by causing stenosis[353; 355], other mutations, however, may have different effects on the artery. Some studies suggested other RNF213 mutations were associated with different types of MMD, i.e., a rare variant p.Ala4399Thr with intermediate frequency in East Asians was associated with hemorrhage type of MMD[356] affected individuals[356], in contrast to that p.Arg4810Lys which was associated with the ischemia type.

Rare variants in RNF213 were found in MMD cases who were not have the p.Arg4810Lys mutation and who were of Japanese [357], Taiwanese[358] and non-Asians[349] ethnicities. Of the 30 rare functional variants found in 649 MMD case-controls in Japan, 16 of these were only observed in affected individuals; additionally, 3 novel variants were found in 31 Taiwanese MMD subjects and 7 rare variants were found in 24 individuals with diverse ethnicity, respectively. Compared to the Japanese case-control study[357], our result has a significantly larger proportion of *RNF213*

mutations in affected individuals compared to controls (Cochran-Mantel-Haenszel Test, $p<0.0001$) (**Figure 15**).

AAA+ ATP domain variants and the risk of IA

One of the very important functions of RNF213 is likely mediated by the two AAA+ regions located in exon 29, which exhibit ATPase activity[336]. In our study focused in IA, several deleterious mutations in the AAA+ domains were found, whereas other studies focused on MMD affected individuals did not (**Figure 16**)[349; 359]. We may assume that variants in different locations of the *RNF213* gene are risk factors for different cerebrovascular disorders by affecting the gene function differently. Cecchi *et al.*[349] suggested the C-terminal domain of RNF213 to be the main risk region for MMD, we believed that exon 29 which contains AAA+ domains may be the risk region for IA.

In our study, we found that a gain of ATPase activity may lead to increased risk of IA. A recent study suggested that the p.Glu2488Gln variant in the 1st AAA+ domain and the p.Arg4810Lys variant reduced protein ATPase activity in the MMD model[342], inhibiting angiogenesis, opposite to our observation. The discordance may lay with the pathological differences between MMD and IA, although both have disrupted vascular walls, the main characteristic of MMD is progressive stenosis accompanied by aberrant angiogenesis of collateral vasculature [360], while IA lesions are characterized by vascular wall weakness and dilatation, accompanied by vascular remodeling and inflammatory response. Actually, the increased angiogenic activity may be a characteristic of the formation of aneurysm lesion[361], and there were studies suggesting high expression of angiogenic factors in IA tissues[362; 363]. Another paper also stated that the ATP concentration is lower in the aneurysm walls than in normal artery tissue[364]. We hypothesize that increased ATPase activity of RNF213 may lead to increased or unbalanced

level of angiogenesis, therefore associate with the formation of aneurysmal lesion. Although the *Rnf213* KO models inflict different angiogenesis alternations under different circumstances[342; 346], we imply that *RNF213* is critical in vascular wall remodeling which can explain the pathogeneses of both MMD and IA. Additionally, the increased ATP hydrolysis may also affect the oxidative phosphorylation process[365], which plays an important role in cases including hypertension[366] and heart failure. Both IA individuals carrying AAA+ domain mutations shown signs of hypertension, one of them were recorded to also suffer multiple cardiovascular problems, including hypercholesterolemia, myocardial infarction and arrhythmia.

RNF213 variants in general populations

We found that an *RNF213* SNP rs6565666 was associated with IA in FCs. Interestingly, when compared to other populations, we found a gradual decrease in the non-ancestral allele (A) frequency, which matches the out-of-Africa migration pattern (**Figure S3**), and may be the result of genetic drift, as this non-ancestral allele also might be a risk allele of IA (dominant model for allele A, P=0.007, corrected). The Finnish have a relatively high frequency of allele A (MAF.FIN=0.26) compared to other European populations (MAF.CEU=0.18); which correlate with their high incidence of IA[89; 213].

When further exploring the *RNF213* variant distributions across populations, we noted that groups of Native American population (Inuit, Argentinian and Peruvian) diverged the most from other populations, which fit the migration patterns. However, if focused on potential functional variants of *RNF213*, East-Asians (Chinese, Japanese and Vietnamese) seem to have the highest diversity (**Figure 13**) beside Native American peoples. The results seem in accord with the fact that *RNF213* East-Asian specific variants are risk factors for MMD, which is also an East-Asian high prevalence disease. The significant difference in *RNF213* burden between FC and East Asians

further suggested the variations in this gene lead to different risk among those two populations. The potential genetic drift of *RNF213* variations may explain the difference between MMD prevalence and RNF213 founder variants in only East Asians but not in other populations. The exploration of population difference of *RNF213* variants also provided additional evidence that the risk gene of IA may be ethnically different[89]. Additionally, as there have been studies suggesting Inuit have a high prevalence of IA, it would be interesting to see if other Native Americans also have this predisposition. Nevertheless, we can also assume that *RNF213* may not be a risk factor for IA among those populations based on our observations.

We could not determine any founder mutation that presented in high frequency in FCs that was significantly associated with IA due to the limiting number of our subjects. The population control used in this study were mostly unscreened for IA, which may also lead to an underlying bias to our results. Only 10% of the affected individuals carry rare, deleterious variants in *RNF213*, suggesting the gene is only a risk factor to some of the cases and that the pathogenesis of IA is very complex and heterogeneous. Additionally, since *RNF213* mutations have not been reported in IA cases of other ethnicities, and that the risk genes of MMD also differs population-wide, suggesting *RNF213* variations may associate with different vascular disorders in different populations. We further assume that different genes are likely to contribute to IA in other populations.

In conclusion, our study supports a role for *RNF213* as a risk factor for IA, possibly via variants affecting different portions of the protein and lead to increased ATPase activity.

ACKNOWLEDGEMENT

P.X. and Z.G.O. are the recipient of fellowship from the Canadian Institutes of Health Research. G.A.R. funded this study from the Canadian Hearth and Stroke Foundation. G.A.R. holds the Canada's Research Chair in Neurogenetics and the Wilder Penfield Chair in Neuroscience. The NeuroX SNP genotyping was partially funded by a grant from Michael J. Fox Foundation to Z.G.O.

Web Resources

IGV, http://www.broadinstitute.org/igv/

NCBI Genome build GRCh37, ftp://gsapubftp-anonymous@ftp.

broadinstitute.org/bundle/2.8/b37/human_g1k_v37.fasta.gz

OMIM, http://www.omim.org

RefSeq, http://www.ncbi.nlm.nih.gov/refseq/

Figures

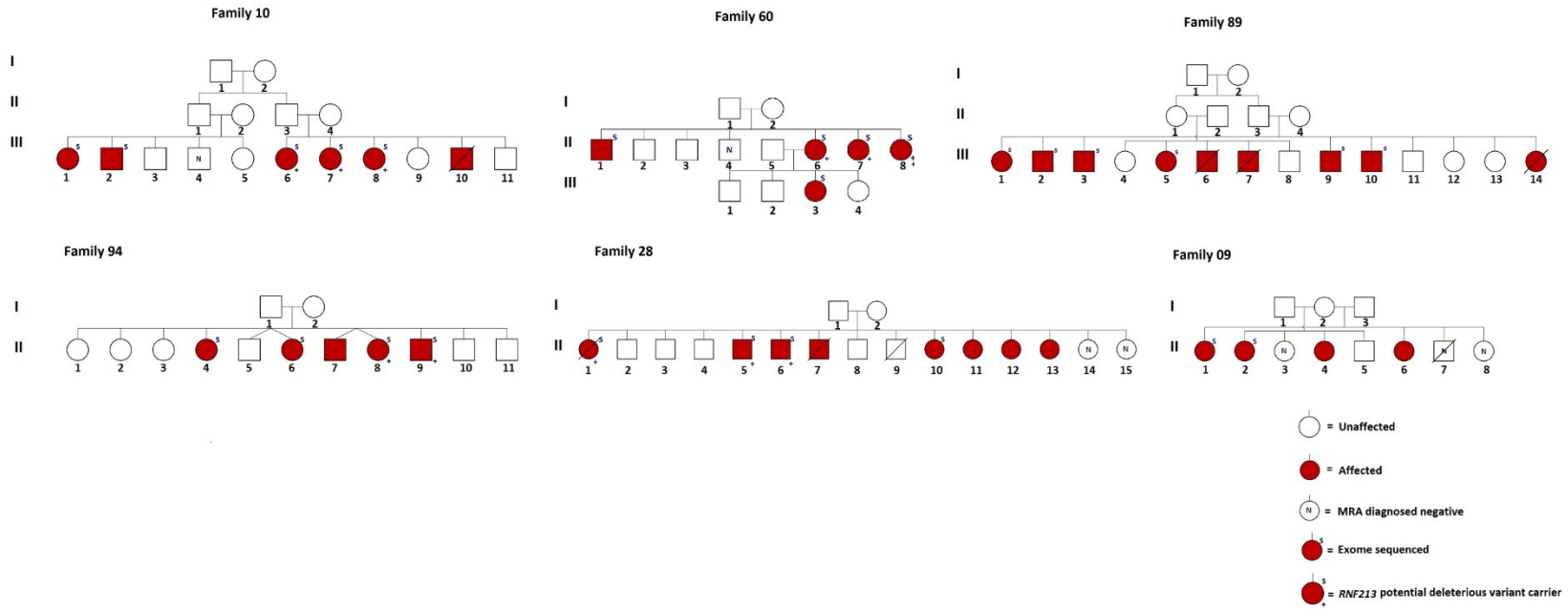

Figure 12. Six IA pedigrees with 26 affected individuals from the initial cohort

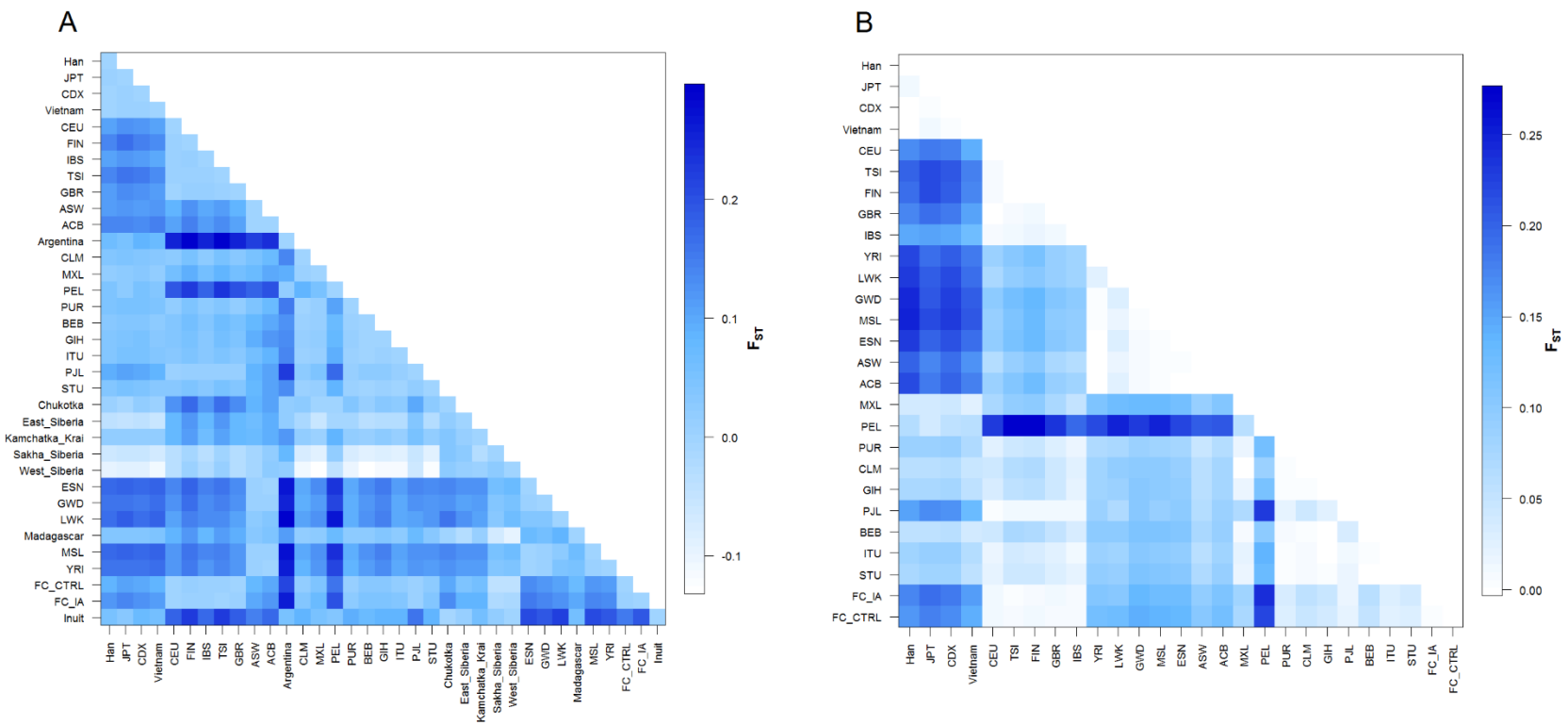

Figure 13. Matrix of pairwise F_{ST} of *RNF213* between world-wide populations

F_{ST} was calculated by *RNF213* common variants (A) from Omni chip and functional variants (B) from exome sequencing between each two populations. FC_CTRL and FC_IA are French-Canadian control cohort and IA cohort, other populations indicated by three letter codes are in accord with 1KGP. The allele distributions of *RNF213* common SNPs varies among populations, suggesting the existence of genetic drift. Among others, there is a big difference in the allelic distributions of RNF213 functional variants between Asians and FC populations.

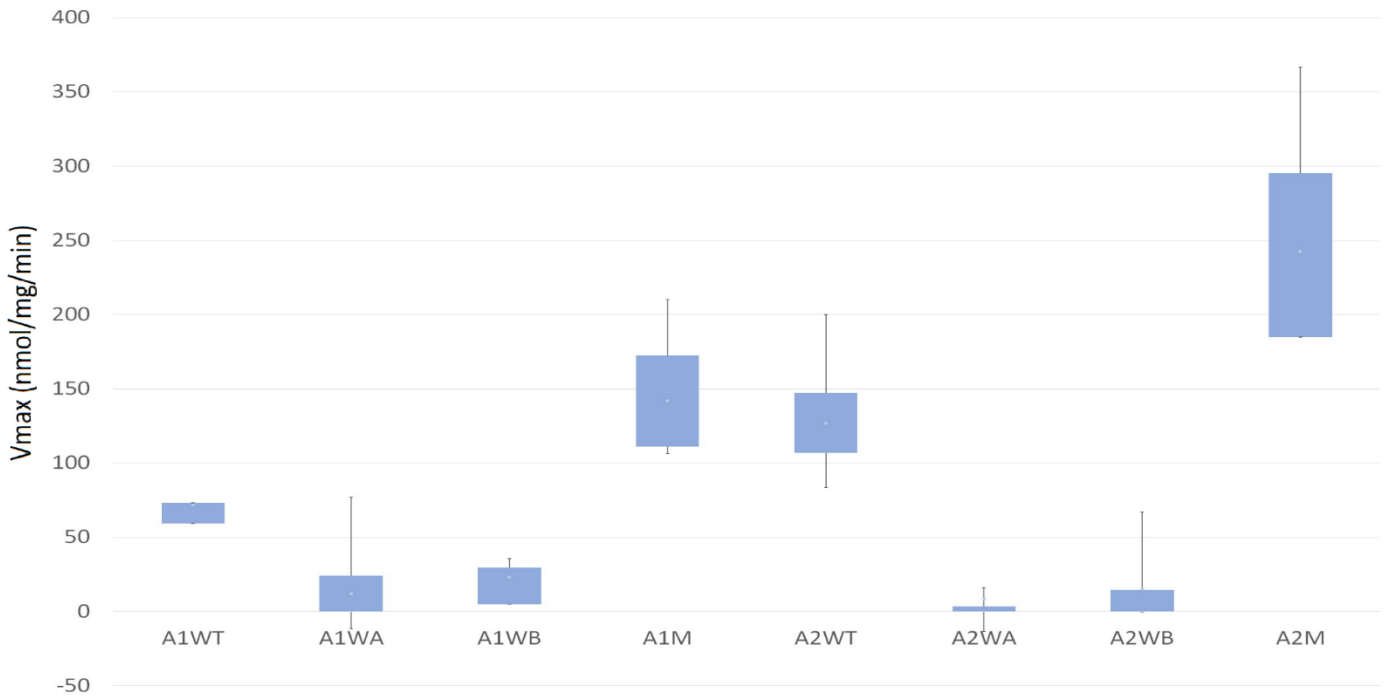

Figure 14. ATPase activity of the wildtypes, variants and LOF controls of the two AAA+ domains of RNF213

Box-and-whisker plot showing ATPase activity measured as nmol of free-phosphate release per mg protein per minute. A1WTand A2WT are the wild-types of the first and second AAA+ modules, respectively; A1M and A2M are the p.Arg2438Cys and p.Ala2826Thr variants, respectively. A1WA, A1WB, A2WA, and A2WB are four LOF controls (p. Lys2426Ala, p. Glu2488Ala, p.Lys2775Ala, and p. Glu2845Ala, respectively) located in the walker A and B domains of the first and second AAA+ modules. The results were generated by eight ATPase assays from two independent cell cultures; each sample is measured in duplicate. Boxes show the interquartile range, and whiskers show maximum and minimum measurements.

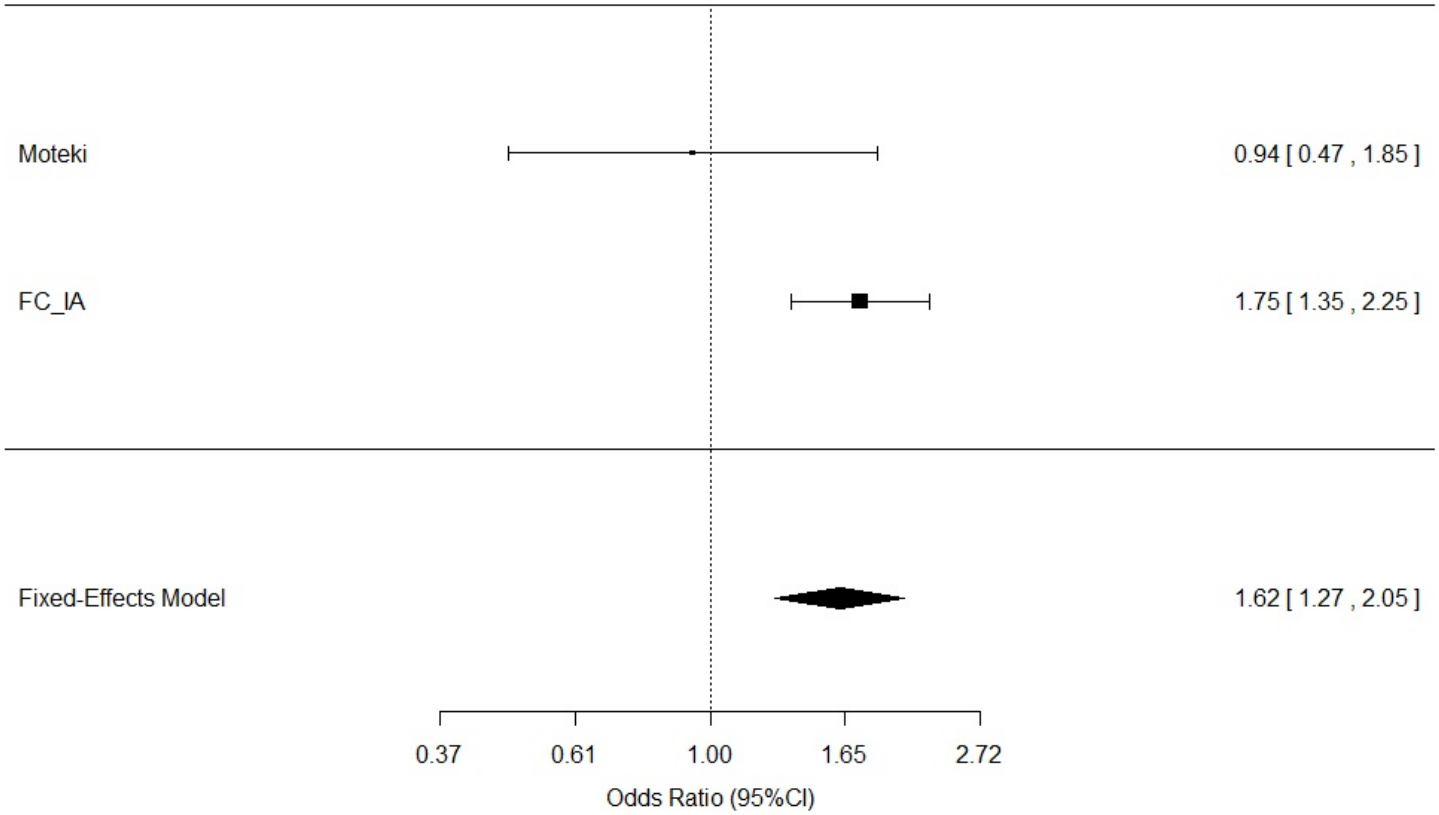

Figure 15. Meta-analysis of two RNF213 studies in different populations

Forest plot of Moteki, Y *et al.* [357] and the current study. Result shown significantly more RNF213 mutations in FC IA affected individuals (Cochran-Mantel-Haenszel Test, $p<0.0001$).

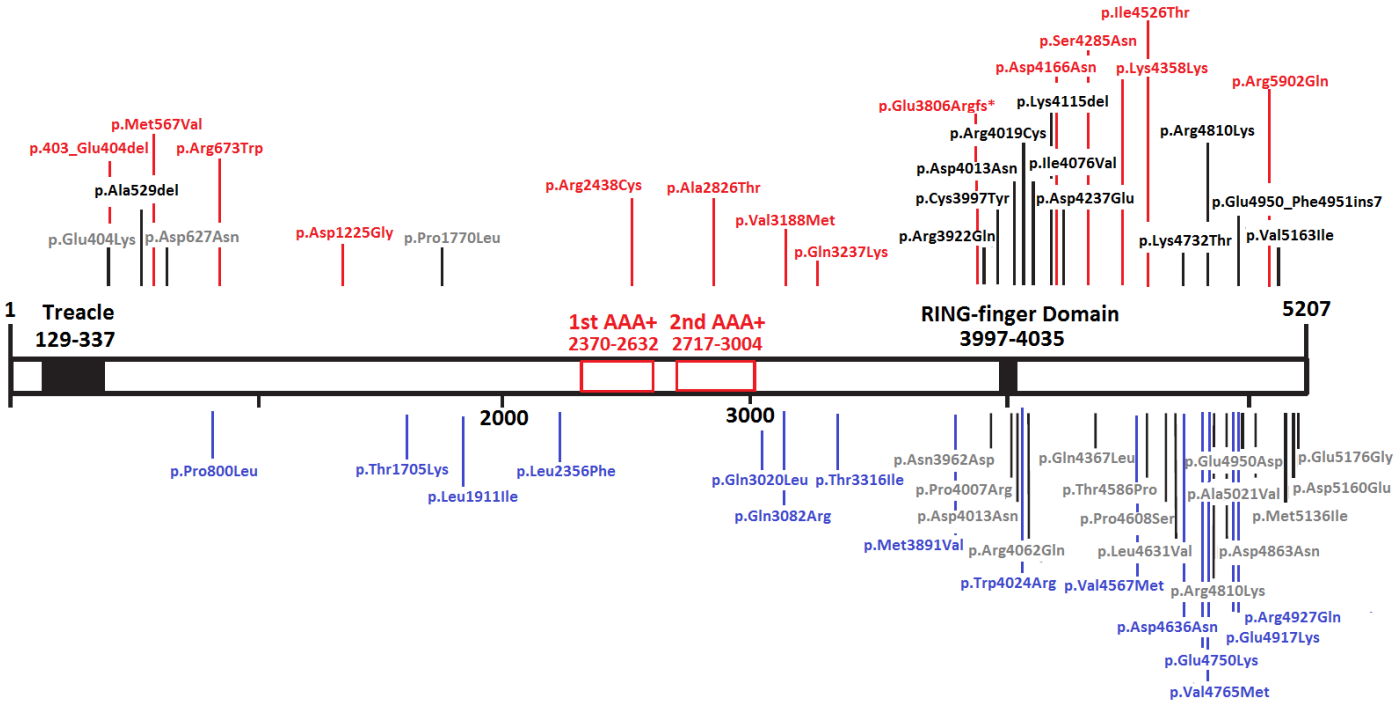

Figure 16. RNF213 missense patient-only variants in IA and MMD

Figure reconstructed based on Cecchi *et al.* (2014)[349] and Koizumi *at al.* (2016)[359]. 14 variants found in the current study are marked in red, variants reported in Moteki et al study are marked in blue, other MMD variants reported in Cecchi et al study are marked in black and the greys are the MMD variants found in other studies. The result showed FC IA affected individuals harbor more deleterious mutations in the AAA+ domains compared to other populations.

Tables

Table VIII. Exome variant filtration steps and results of initial cohort

Coding and splicing variants after QC	**Family 60**	**Family 89**	**Family 10**	**Family 09**	**Family 28**	**Family 94**
A:Variants in 1+ family affected	33,061	38,609	37,706	31,975	33,090	32,727
B:functional variants	18,455	21,857	21,255	17,602	18,417	18,118
C:French-Canadian specific variants	1,489	2,041	1,797	825	1,089	882
D:exclude most variable exonic genes	999	1,388	1,157	595	670	627
E:inhouse database <5	403	581	509	309	360	294
F:Variants in 2+ family affected	178	258	226	50	236	201
G:deleterious(PP2>0.8 or GERP>2, and frameshift)	79	109	103	15	99	67
Number of genes	78	108	102	15	90	64
Gene with cross-family recurrent variants (variants)	6 (6)	7 (7)	6 (6)	2 (2)	8 (9)	11 (13)
Genes with cross-family recurrent variants (variants)	19 (35)					
Gene with different variants	11					
Gene with 2+ different variants	1 (*RNF213*)[a]					

a. p=0.01, VT test, bonferroni correction, 5000 permutations.

Table IX. Prioritized genes with French-Canadian specific deleterious variants

Gene	**Chr**	**Position**	**Annotation**	**Family ID**	**PP2**	**GERP++**
ABCA10	17	67215903	NM_080282:c.313A>G;NP_525021:p.Asn105Asp	60	0.995	1.83
	17	67221498	NM_080282:c.2T>C; NP_525021:p.Met1Thr	94	0.998	3.6
AIM1	6	107008779	NM_001624:c.4733G>T; NP_001615:p.Gly1578Val	89	1	3.4
	6	106968391	NM_001624:c.2084C>A; NP_001615:p.Ser695Tyr	94	0.966	5.62
CDAN1	15	43028521	NM_138477:c.548C>T; NP_612486:p.Ser183Leu	89	0.127	3.91
	15	43023250	NM_138477:c.1880C>T;	28	0.007	5.49

			NP_612486:p.Ala627Val			
GPATCH8	17	42477055	NM_001002909:c.2390G>A; NP_001002909:p.Arg797Gln	28	0.658	2.67
	17	42475271	NM_001002909:c.4174G>A; NP_001002909:p.Gly1392Ser	10	0.98	4.57
HELZ2	20	62196325	NM_033405:c.2143C>T; NP_208384:p.Arg715Trp	28	0.858	4.11
	20	62195122	NM_033405:c.3346C>T; NP_208384:p.Arg1116Trp	60	1	-0.842
RNF213	17	78272125	NM_001256071:c.2017C>T; NP_001243000:p.Arg673Trp	28	0.997	0.06
	17	78319115	NM_001256071:c.6980A>G; NP_001243000:p.Asn2327Ser	60	0.006	2.19
	17	78293222	NM_020954.3:c.3134C>T; NP_066005.2:p.Ser1045Leu	10	0.496	2.628
	17	78353451	NM_001256071:c.13577T>C; NP_001243000:p.Ile4526Thr	94	0.291	2.81
	17	78336958	NM_001256071:c.11413del; NP_001243000: p.Glu3806Argfs*27	60		
OR11H1	22	16449405	NM_001005239:c.400G>A; NP_001005239:p.Ala134Thr	89	0.852	0.664
	22	16449399	NM_001005239:c.406G>A; NP_001005239:p.Asp136Asn	10	0.988	1.84
PLEC	8	144998555	NM_201384:c.5542C>T; NP_958786:p.Leu1848Phe	60	0.726	4.06
	8	144996050	NM_201384:c.7939G>T; NP_958786:p.Asp2647Tyr	89	0.992	0.966
RTTN	18	67806837	NM_173630:c.2786G>A; NP_775901:p.Arg929Lys	94	0.985	4.96
	18	67742698	NM_173630:c.4454A>G; NP_775901:p.His1485Arg	89	0.043	2.92
SF3A2	19	2248232	NM_007165:c.1082C>G; NP_009096:p.Ala361Gly	28	0.474	2.57
	19	2245459	NM_007165:c.260C>T; NP_009096:p.Ala87Val	94	0.024	4.54
ZNF335	20	44577719	NM_022095:c.3902C>T;	60	0.201	4.91

			NP_071378:p.Ala1301Val			
	20	44596594	NM_022095:c.593A>G; NP_071378:p.Asp198Gly	10	0.977	3.92

Table X. Mutation burden of 11 prioritized genes between FC IA patients and controls

Gene	Risk/total allele[a] in affected individuals	risk/total allele in controls	VT statistic	p (unadj)	std error	permutations
ABCA10	2/36	2/212	0.709677	0.0775845	0.347336	5,000
AIM1	2/36	3/212	0.569837	0.126873	0.339132	1,000
CDAN1	2/36	2/212	0.709677	0.0771846	0.346911	5,000
GPATCH8	2/36	1/212	0.903274	0.0259948	0.345287	5,000
HELZ2	2/36	4/212	0.569837	0.135864	0.340307	1,000
OR11H1	2/36	3/212	1.20892	0.0195961	0.398189	5,000
PLEC	2/36	15/212	0.121577	0.532468	0.376399	1,000
RNF213	5/36	4/212	1.50576	0.00119976[b]	0.354706	5,000
RTTN	2/36	5/212	0.371868	0.260739	0.344665	1,000
SF3A2	2/36	2/212	0.903274	0.0313937	0.325226	5,000
ZNF335	2/36	1/212	0.903274	0.0269946	0.378224	5,000

a. Risk alleles were alleles harboring FC specific deleterious variants, total alleles were calculated by the founder alleles of 6 IA families.
b. Corrected p=0.013.

Table XI. Patient-only and control-only *RNF213* variants which predicted deleterious

Position	Annotations	affected individuals (233)	Control individuals (277)	PP2	GERP++	Freq in EVS
78264463	c.1208_1210del:p.403_404del	1	0	1	0	0.001917
78268746	c.1699A>G:p.Met567Val	1	0	0.999	2.72	0
78272125	c.2017C>T:p.Arg673Trp	3	0	0.997	0.06	0
78305962	c.3674A>G:p.Asp1225Gly	1	0	0.919	0	0
78319447	c.7312C>T:p.Arg2438Cys	1	0	0.655	2.35	0.000077
78320611	c.8476G>A:p.Ala2826Thr	1	0	1	5.59	0
78321697	c.9562G>A:p.Val3188Met	1	0	0.96	5.36	0

78321844	c.9709C>A:p.Gln3237Lys	3	0	0.264	2.61	0
78345744	c.12496G>A:p.Asp4166Asn	1	0	1	5.02	0.000923
78346877	c.12854G>A:p.Ser4285Asn	1	0	0.024	2.7	0.000077
78348389	c.13074G>A:p.Lys4358Lys	1	0	1	0	0
78353451	c.13577T>C:p.Ile4526Thr	3	0	0.290848	2.81	0
78363707	c.15275G>A:p.Arg5092Gln	1	0	0.998	3.14	0.000077
78321941	c.9806G>A:p.Arg3269Gln	0	2	1	3.02	0.000077
78328364	c.10850C>T:p.Ala3617Val	0	1	0.555	4.79	0
78291060	c.2884G>A:p.Glu962Lys	0	1	0.068	3.66	0
78321560	c.9425T>C:p.Val3142Ala	0	1	0.998	5.09	0

Table XII. *RNF213* variable thresholds of world-wide populations compared to French-Canadians

	sample	**# variants**	**Total mac**	**statistic**	**p-value**	**SE**	**permutations**
CHB	380	55	2,333	3.95287	0.00019996	0.364289	5,000
JPT	381	54	2,453	1.79685	0.00119976	0.346035	5,000
CDX	370	54	2,306	3.21125	0.00019996	0.343601	5,000
KHV	376	51	2,322	3.61053	0.00019996	0.355326	5,000
CEU	376	54	2,187	0.955313	0.113629	0.346662	3,000
TSI	384	54	2,151	1.23624	0.0319936	0.364394	5,000
GBR	368	52	2,108	1.0242	0.0657868	0.349268	5,000
FIN	376	51	2,104	1.48321	0.00679864	0.358871	5,000
IBS	384	57	2,250	1.01387	0.124875	0.388177	1,000
YRI	385	83	2,666	0.741073	0.361638	0.421002	1,000
LWK	376	80	2,545	-0.0442278	0.936064	0.415337	1,000
GWD	390	81	2,763	-0.0641889	0.924076	0.433297	1,000
MSL	362	82	2440	-0.177864	0.973027	0.401114	1,000
ESN	376	84	2,598	-0.6186	0.998002	0.436067	1,000
ASW	338	82	2,173	-1.80752	1	0.39357	1,000
ACB	373	82	2,552	-0.38227	0.99001	0.43374	1,000
MXL	341	55	2,060	0.660489	0.275724	0.302976	1,000
PUR	381	66	2,335	-0.359017	0.995005	0.379015	1,000
CLM	371	69	2,293	-0.673233	1	0.420994	1,000
PEL	362	58	2,313	1.48289	0.0039992	0.350589	5,000

GIH	380	53	2,312	1.47606	0.00559888	0.355099	5,000
PJL	373	56	2,169	1.07995	0.0631874	0.351286	5,000
BEB	363	58	2,212	0.293378	0.743257	0.33014	1,000
STU	379	56	2,278	0.959633	0.135864	0.362751	1,000
ITU	379	57	2,271	0.434712	0.601399	0.349177	1,000
FC_IA	513	72	2,938	1.69234	0.0125975	0.454633	5,000

Abbreviations are as follows: MAC, minor allele count; CHB, Han Chinese in Beijing, China; JPT, Japanese in Tokyo, Japan; CDX, Chinese Dai in Xishuangbanna China; KHV, Kinh in Ho Chi Minh City, Vietnam; CEU, Utah residents with northern and western European ancestry from the CEPH collection; TSI, Toscani in Italia; GBR, British in England and Scotland; FIN, Finnish in Finland; IBS, Iberian population in Spain; YRI, Yoruba in Ibadan, Nigeria; LWK, Luhya in Webuye, Kenya; GWD, Gambian in western divisions in the Gambia; MSL, Mende in Sierra Leon; ESN, Esan in Nigeria; ASW, Americans of African ancestry in southwest USA; ACB, African Caribbeans in Barbados; MXL, Mexican ancestry from Los Angeles, USA; PUR, Puerto Ricans from Puerto Rico; CLM, Colombians from Medellin, Colombia; PEL, Peruvians from Lima, Peru; GIH, Gujarati Indian from Houston, Texas; PJL, Punjabi from Lahore, Pakistan; BEB, Bengali from Bangladesh; STU, Sri Lankan Tamil from the UK; ITU, Indian Telugu from the UK; and FC_IA, French-Canadian individuals with intracranial aneurysms.

3.3: GENOME-WIDE ASSOCIATION ANALYSIS IDENTIFIES NEW CANDIDATE RISK LOCI FOR FAMILIAL INTRACRANIAL ANEURYSM IN FRENCH-CANADIANS

Manuscript: Genome-wide association analysis identifies new candidate risk loci for familial intracranial aneurysm in the French-Canadian population

Authors and affiliations:

Sirui Zhou[1,2], Ziv Gan-Or[1,4], Amirthagowri Ambalavanan[1,4], Pingxing Xie[1], Cynthia V. Bourassa[1], Stephanie Strong[1,4], Jay P. Ross[1,4], Alexandre Dionne-Laporte[1], Dan Spiegelman[1], Nicolas Dupré[5], Lan Xiong[2,3], Patrick A. Dion[1,6], Guy A. Rouleau[1,6,*]

1) Montreal Neurological Institute and Hospital, McGill University, Montréal (QC), Canada; 2) Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal (QC), Canada; 3) Centre de recherche, Institut universitaire en santé mentale de Montréal (QC) Canada; 4) Department of Human Genetics, McGill University, Montréal (QC) Canada; 5) Faculty of Medicine, Université Laval, Quebec (QC), Canada; 6) Department of Neurology and Neurosurgery, McGill University, Montréal (QC) Canada.

Authors contribution:

Study design: SZ, GAR, PAD, LX

Experiment: SZ, SS, JPR

Bioinformatic analysis: SZ, AA, PX, ADL, DS

Sample recruitment: CVB, ZGO, ND

Manuscript writing: SZ, LX, PAD, GAR

ABSTRACT

Intracranial Aneurysm (IA) is a common disease with a worldwide prevalence of 1-3%. In French-Canadian (FC) population, where there is an important founder effect, the incidence of IA is higher and is frequently seen in families. In this study, we genotyped a cohort of 257 FC familial IA patients and 1,992 FC controls using the Illumnia NeuroX SNP-chip. The most strongly associated loci were tested in 34 Inuit IA families and in 32 FC IA patients and 106 FC controls that had been exome sequenced. After imputation, one locus at 3p14.2 (*FHIT*, rs1554600, $p=4.66\times10^{-9}$) reached a genome-wide significant level of association and a subsequent validation in Nunavik Inuit cohort further confirmed the significance of the *FHIT* variant association (rs780365, FBAT-O, p=0.002839). Additionally, among the other 28 promising loci ($p<5\times10^{-6}$), the one at 3q13.2 (rs78125721, $p=4.77\times10^{-7}$), which encompasses *CCDC80,* also showed an increased mutation burden in the WES data (*CCDC80*, SKAT-O, p=0.0005). The product of *CCDC80* has been reported to have an important function in angiogenesis. In this study, we identified two new potential IA loci in the FC population: *FHIT,* which is significantly associated with hypertensive IA, and *CCDC80,* which has potential genetic and functional relevance to IA pathogenesis. Our results also suggest that the genetic background of familial IA might be different from the risk loci revealed by previous IA GWA studies and that population-wide genetic heterogeneity may also be an important variable for IA.

INTRODUCTION

Intracranial Aneurysm (IA) has a prevalence of 1-3% in the general population[102; 312]. The rupture of an IA can lead to subarachnoid hemorrhages (SAH), which has devastating consequences. Environmental and genetic factors, such as hypertension and smoking[109], family history and ethnicity all contribute to the risk of IA. Because of the complexity of IA, genome-wide association studies (GWAS) have become the predominant strategy lately used to look for genetic factors associated with IA. These studies used several large cohorts with IA patients, mainly of Finnish, Japanese or European descent. Several risk loci were discovered in these GWA studies: 8q11.23 (*SOX17*), 9p21.3-23.1 (*CDKN2A-CDKN2BAS*)[207; 214] and 2q33.1 from the European and Japanese cohorts; 18q11.2 (*RBBP8*), 13q13.1 (*STARD13-KL*) and 10q24.32[208] from the Finnish and Japanese cohorts; 1q23.1, 3p25.2, 7p21.2, 9q31.3[208] and 4q31.22 (*EDNRA*)[211] from two Japanese cohorts, and 7p21.1 (*HDAC9*)[215] from a cohort with European ancestry. A more recent Finnish IA study revealed additional GWAS risk loci, including 2q23.3, 5q31.3 and 6q24.2, represented by low-frequency SNPs[89]. Multiple GWAS signals suggest that the genetic etiology of IA may be complex and population specific. The risk loci found in each GWA study was estimated to only account for 4.1%-6.1% of the heritability in the respective cohort[89].

It has been reported that French-Canadian (FC) population has higher IA/SAH incidence and that patients usually aggregate in large pedigrees[321], with 30% of IA patients having a family history (fIA)[322]. Similarly, to the Finnish people, French-Canadians are also descended from a relatively small founder population and have population specific variants due to the population bottleneck and genetic drift. Therefore, we hypothesized that population specific and median/low frequency variants may play an important part in the disease risk in FC fIA.

RESULTS

GWAS discovery phase

After data QC and sample pruning, 621,983 SNPs with 173 FC IA cases and 1,772 FC controls remained in the analysis. The genome-wide threshold for significance after Bonferroni correction was set to 5×10^{-8}. Marker-wise P values of Cochran–Armitage trend test were performed using PLINK 1.9[273] for genotyped variants. Genomic inflation factor $\lambda = 1.02$ indicated that there was little inflation of excessive significant markers, as shown in quantile-quantile (QQ) plot (**Figure S1**).

The result of the initial trend test showed 3q13.2 as the most significant locus: rs2705520 ($p=6.93\times10^{-8}$), an intronic SNP in *ATG3* (*autophagy 3*), followed by rs1877362 ($p=1.16\times10^{-7}$) in *CCDC80* (*coiled-coil domain containing 80*) and rs1472107 ($p=1.18\times10^{-7}$) in *SLC35A5* (*Solute Carrier Family 35 member A5*) (**Figure S2**).

After imputation using the Haplotype Reference Consortium (HRC) and the exclusion of low quality and low MAF variants, 7,614,484 remaining variants were included in the test of associations using SNPtest[367]. The results were shown in the Manhattan plot (**Figure 17**). One locus reached genome-wide significant level after imputation: 3p14.2 (rs1554600, OR 0.26, $p=4.66\times10^{-9}$) located in gene *FHIT*. TaqMan validation of rs1554600 on IA cases and controls suggested the imputation was accurate (MAF.case=0.0838 and MAF.control=0.026). 28 most significant SNPs, representing distinct loci that each reached promising level ($p<5\times10^{-6}$), were prioritized for further validation (**Table XIII**).

Using imputed SNPs, the two loci 3p14.2 and 3q13.2 were estimated by GCTA-GREML[368-370] (Genome-wide Complex Trait Analysis) to account for approximately 3% of the heritability in the FC cohort (standard error (SE) =0.026).

Replication in exome data and Inuit cohort

We looked into the exome sequencing data of the aforementioned 28 genes containing the top GWAS SNPs in our FC WES cohort; of the 25 genes that had exonic variants, a total of 186 exonic and splicing variants were found in 138 FC cases and controls. Sequence Kernel Association Test (SKAT)[371] results showed excessive exonic variation burden in IA cases in four genes *SLC35F3* (p=0.002), *DTNB* (p=0.003), *CCDC80* (p=0.0005) and *PABPC3* (p=0.0001) (**Table XIV**). However, the first two genes were less convincing with the limited number of variants in the testing (two variants for each). *PABPC3* was unlikely to be a risk gene for IA due to its human testis-specific expression. Thresholds Test (VT)[325] focused on the selected genes revealed that only *CCDC80* (p=0.01) in 3q13.2 reached the statistical significance after accounting for multiple testing.

In the Nunavik Inuit IA cohort, of the 207 SNPs within the 28 aforementioned genes and neighboring regions which were found to have $p<0.05$ in Family Based Association Test (FBAT)[372], 50 SNPs located in the *FHIT* gene region showed significant associations (top SNP rs780365, p=0.002839) (**Table S1**). Although the associations were no longer significant after corrections of multiple testing, the similar results established in two distinct populations still suggested that intronic variants in *FHIT* may likely be associated with IA.

Replication of previous GWAS risk loci

A regional Manhattan plot covering the 13 previously identified GWAS risk loci across our FC IA cohort is shown in **Figure S3**, where the top association signal was located in the 4q region near *EDNRA* (endothelin receptor type A) (rs114324350, $p=4.8\times10^{-5}$). The two IA risk loci 9p (*CDKN2BAS*) and 8q (*SOX17*), which were discovered and replicated at least once in previous GWAS [207; 214], showed significant associations but did not reach genome-wide significant levels (top signal: rs575427, *CDKN2BA*S, $p=1.2\times10^{-3}$ and rs4738913, *SOX17*, $p=2.3\times10^{-3}$; previous SNP: rs1333040, p=0.008565 and rs10958409, p=0.006042).

Overall these 13 previously defined loci encompassed 68 genes for which our FC IA exomes revealed 497 exonic variants (**Table S2**). Among those genes only *STARD13* (star-related lipid transfer (START) domain containing 13) showed a significant variation burden in cases compared to control individuals (VT test, $P = 0.0476$, adjusted). While *STARD13* was previously associated with IA[208], it was not replicated in our GWAS data (top signal: rs9536854, $P = 1.6 \times 10^{-3}$).

DISCUSSION

In this study, we discovered a new IA associated region on 3p14.2 which encompasses gene *FHIT* in the French-Canadians (**Figure 18**), intronic variants in *FHIT* also showed significant association with IA in the Nunavik Inuit population. Additionally, we found evidence suggesting exonic variants in *CCDC80* within the 3q13.2 locus to be associated with the French-Canadian IA cases. Collectively, SNPs in *FHIT* and *CCDC80* could explain approximately 3% of the heritability of IA in French-Canadians, higher than that was reported in the Finnish study (2.1%)[89] and in the GWAS replication (2.5%)[208]. The underlying reason for this might be that French-Canadians are

a more homogenous population and this study has mainly included familial cases. On the other hand, two of the previously identified GWAS loci, encompassing the genes *CDKN2BAS* and *SOX17*, showed only a moderate association in the French-Canadian founder population.

Although the number of cases in this study was limited, we tried improving the power by doing the following: 1) targeting IA patients with family history, 2) focusing on individuals only with the French-Canadian ethnicity, and 3) validating the findings in exome sequencing results and in another founder population. Because French-Canadian originated from a small founder population, we also included intermediate variants after the imputation. The top SNPs that we discovered in *FHIT* and *CCDC80* had rare or intermediate frequency (2-4%). rs1554600 in *FHIT* has a slightly higher frequency in French-Canadians (3.3%) compared to 1KGP CEU population (1.9%), which suggest a bottleneck drift may be the reason for the accumulation of low-frequency variants that potentially associated with the risk of IA in the French-Canadians.

FHIT (fragile histidine triad) is a tumor suppressor gene that regulates DNA replication and signals stress responses[373], and encompasses the most active of the common human chromosomal fragile regions (FRA3B); and its expression has an important role in response to oxidative damage[374]. On the other hand, oxidative stress is known to be a key contributor to IA formation and rupture[375]. Interestingly, a study highlighted that the SNPs in *FHIT* have been associated with hypertensive traits in populations from Saguenay-Lac-St-Jean region, mainly French-Canadians[376]. As over 40% of our French-Canadian IA cases are also affected with hypertension (**Table XVI**), we considered that rs1554600 in *FHIT* is more likely to be a risk for hypertensive IA in French-Canadians. Further test of *FHIT* SNPs between IA patients with and without hypertension revealed several SNPs in LD to be significantly associated with this trait (rs73098963, p=0.002611, GWAS p value=2.42×10^{-8}).

The 3q13.2 locus is a gene-rich region with many function in inflammatory responses. *BTLA* (B- and T-lymphocyte attenuator) is involved in inflammatory responses[377] and homeostasis of the immune system[378]. A previous study showed *BTLA* expression to be up-regulated in organs after a hemorrhagic shock[379]. The 3q13.2 locus also comprises *ATG3,* which encodes a protein known to induce apoptosis[380] and also to act as a regulator of oxidant and inflammatory balance that regulates endothelial cell stress response[381]. The top SNP rs2705520 in *ATG3* was also reported in a GWAS to be associated with asthma[382], suggesting its role in inflammatory diseases.

The most interesting gene in the 3q13.2 locus is *CCDC80*, also known as *SSG1* (steroid-sensitive gene 1), which is a cGMP signalling effector and was reported to be widely express in vascular smooth muscle cells[383]. A previous study also showed that it has a role as a modulator of glucose and energy homeostasis[384]. Another study highlighted that the product of fibroblast growth factor (*FGF*) regulates the expression of *CCDC80*[385], which is in turn also involved in cell adhesion during differentiation of fibroblast[386]. *CCDC80* was reported as a tumor suppressor as well[387]. These evidences suggest the critical function of *CCDC80* in vascular formation.

Although no variant from these four regions has shown effects of genetic drift or selection, *CCDC80* does harbor a large number of rare variants in the French-Canadians (**Table XV**); the gene also has significant difference in the variation burden in cases and controls. The 3q13.2 region contains dosage sensitive genes, among which are *ATG3* and *CCDC80*[388; 389], therefore the potential different expression levels of those genes may affect the risk of IA. Interestingly, these two genes were the only ones that have been validated in our Nunavik Inuit cohort, which has a completely different genetic background compared to the French-Canadians.

We have also found evidence of associations in two of the seven genes reported in the previous GWA studies, *STARD13* and *EDNRA*. These genes showed strongest association in an

earlier Finnish study[208; 210]. Previous GWAS risk loci on 9p21 and 8q11 showed weak associations in our cohort. The result is expected as familial IA may have different underlying genetic background then sporadic IA, which were the main recruits of the previous GWAS. In the GWAS conducted by Foroud *et al.*[214], the discovery cohort comprised of familial IA and also suggested different risk loci on 2q32 (*PDE1A*, $p=6.66\times10^{-7}$) and 10q26 (*BTBD16*, $p=6.69\times10^{-6}$).

In conclusion, we have provided evidence for four new loci associated with IA in French-Canadian IA cohort recruited from Montréal and Québec city, which could explain 3% of the disease heritability. Based on the findings of this study and the functions of their encoded products, two genes (*FHIT* and *CCDC80*) are potentially relevant to IA with strong aggregation of familial IA cases with high blood pressure in the French-Canadian population. *FHIT* is more particularly associated with hypertensive IA cases and this may be the consequence of a bottleneck drift that affected the French-Canadian founder population. *CCDC80* was shown to have a large number of rare variations in the French-Canadian cohort and with significantly different variation burdens between IA cases and controls. The lack of association of SNPs in *FHIT* and *CCDC80* in previous GWAS studies suggested a genetic heterogeneity in IA, and thus additional studies targeting other high-risk populations are needed. However, the limited number of cases available in our study warrants a validation study that will have access to a larger cohort from the same founder population, thus to increase the power of detection.

MATERIALS AND METHODS

Discovery cohort

The discovery cohort included 257 French-Canadian IA patients with family history and recruited in Montréal and Québec City, Canada. The diagnoses were confirmed either by magnetic resonance angiography (MRA), or by surgical confirmation (clipped or coiled). An additional 1,992 FCs, comprised of unrelated individuals without cerebrovascular diseases were included as population controls. Their demographic information is listed in **Table XVI**. Written consent was obtained from all individuals and this study was approved by appropriate institutional ethics committees (Québec, Canada).

Genotyping and quality control

All patients and controls were genotyped using the Illumnia NeuroX SNP-chip, which contains 719,885 markers and is comprised of the backbone of Illumnia HumanOmniExpress-v24 BeadChip. Raw data was processed by Illumnia GenomeStudio software before the genotypes were generated. Both markers and samples were passed through a series of quality control (QC) steps. Samples were first removed if duplicated or if they had one of the following issues: 1) sex discrepancies; 2) exceeded a missing rate of 0.02; 3) ethnical admixture determined by PCA; or 4) with cryptic relateness determined by PLINK. Markers were removed if they meet one of the following criteria: 1) exceeding a missing rate of 0.02; 2) having a minor allele frequency (MAF) lower than 0.01; 3) deviating from Hardy Weinberg Equilibrium ($p<0.0001$).

Principal Component Analysis (PCA) impelemented in the package EIGENSOFT 6.0[329] was performed to assess the ethnicity of the samples. Three distinct populations CEU, CHB and YRI from 1000 Genome (1KGP) Phase III were used for clustering and CEU outliners were removed from further analysis. The remaining homogeneous population was also adjusted for the principal components in the subsequent tests for associations.

Imputation

Imputation was done by the Sanger Imputation Server (https://imputation.sanger.ac.uk/) using Haplotype Reference Consortium r1.1[390], and were pre-phased using SHAPEIT2[391]. Imputed variants were included in the further analyses only with MAF > 0.01 and with imputation quality score > 0.3.

Association analysis

Frequentist additive association implemented in SNPtest[367] was used to test for association of the imputed dataset, between FC IA cases and controls. Five major principal components were used as covariates for ancestry adjustment along with the sex of samples. Only autosomal SNPs were analyzed.

Regional association of suggestive loci were plotted using LocusZoom[392] with LD data from 1KGP CEU population.

Heritability estimation

We estimated the heritability from the original and imputed variants within the most promising loci, using methods of Estimation of Variance explained by SNPs (GREML)[368] and GREML-LDMS[370] programs implemented in the package of the Genome-wide Complex Trait Analysis (GCTA)[369].

FC and Inuit IA WES data

We examined the loci from the GWAS signals that have reached suggestive significance ($p<5\times10^{-6}$) in the exome sequencing results of 32 selected FC IA cases and 106 FC controls. Variable Thresholds Test (VT)[325] implemented in Variant Tools (Vtools)[393] and Sequence Kernel

Association Test (SKAT)[371] were performed to test the exonic variation burden in the genes located in the GWAS significant regions.

Thirty-four Nunavik Inuit (Québec, Canada) families comprised of 49 IA patients and 124 family controls were also used to follow up the regions of significance. The samples was genotyped on Illumina HumanOmniExpress-v24 Beadchip which contains 730,525 SNPs. We looked into all the suggestive loci of the FC IA discovery cohort, and performed family-based association analysis (fbat) implemented in FBAT package[372] in the Inuit SNP-chip data to test the case-control association in related individuals.

Replication of previous IA GWAS loci

At this point 13 loci from previous GWA studies (**Table S2**) were selected to examine if they could be replicated in our study. A 2Mb (1Mb each upstream/downstream) region was selected for each previous associated SNP in the data of both FC discovery cohort and FC exome replication cohort. PLINK and Genome Analysis Toolkit (GATK)[267] were used to extract the variants in the two respective datasets.

ACKNOWLEDGEMENTS

P.X. and Z.G.O are the recipient of fellowship from the Canadian Institutes of Health Research. G.A.R. funded this study from the Canadian Hearth and Stroke Foundation. G.A.R. holds the Canada's Research Chair in Neurogenetics and the Wilder Penfield Chair in Neuroscience.

Figures

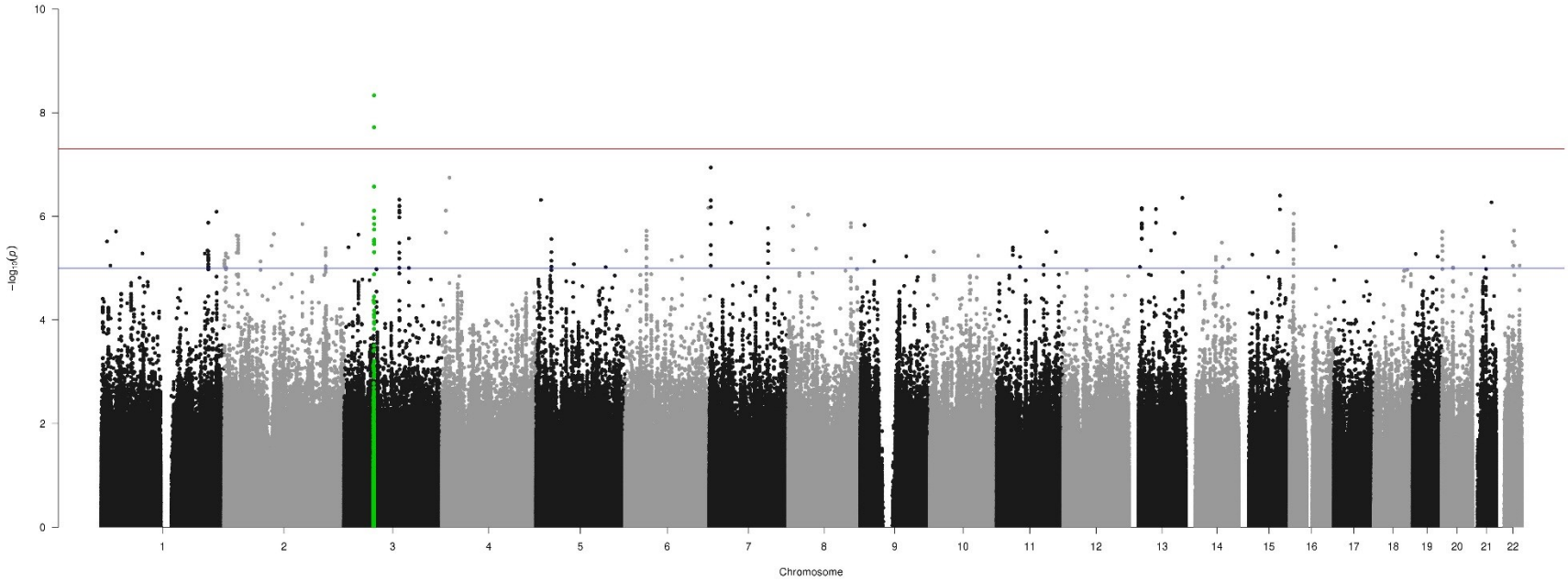

Figure 17. Manhattan plot of FC IA case-control association analysis using imputed data.

Red line indicates the level of genome-wide significant association ($p=5\times10^{-8}$); blue line indicates the level of suggestive association ($p=5\times10^{-6}$). Green dots indicate *FHIT* SNPs.

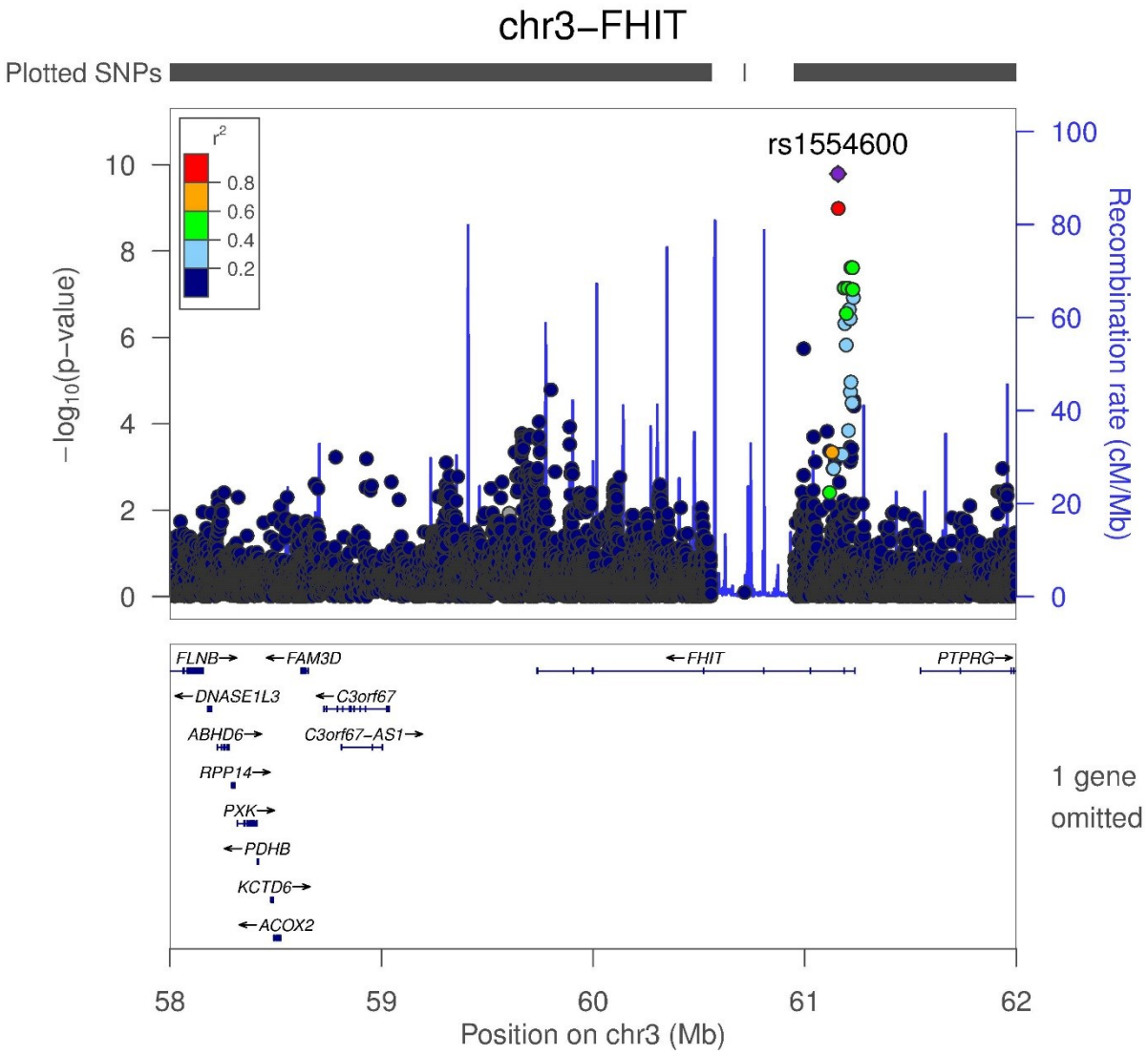

Figure 18. Regional association signals of 3p14.2 locus

4Mb region around the most significant association, rs1554600 in 3p14.2 locus is displayed using imputed data. Purple line indicates the genetic recombination rate (cM/Mb). Variants in linkage disequilibrium is shown in color gradient indicating r2 levels (hg19, 1KGP, Nov 2014, EUR). Variant with the most significant association (rs1554600) is indicated in purple diamond.

Tables

Table XIII. 28 genes with top SNPs reached promising level of association

loci	top SNP	Position (hg19)	Gene (or nearby)	Frequency in cases	Frequency in controls	OR	P-value	INFO
3p14.2	rs1554600	61157774	*FHIT*	0.080925	0.022291	0.258936	4.66E-09	0.94
7p22.2	rs12535623	2951412	*CARD11*	0.260116	0.154345	0.519156	1.15E-07	0.95
4p15.32	rs116130729	16012786	*PROM1*	0.072254	0.023138	0.304125	1.8E-07	0.86
3q13.2	rs78125721	1.12E+08	*CCDC80*	0.043353	0.009029	0.201063	4.77E-07	0.97
21q22.13	rs111610752	38151254	*HLCS*	0.052023	0.016366	0.303181	5.39E-07	0.68
8p23.1	rs117537300	10743939	*XKR6*	0.118497	0.052765	0.414387	6.64E-07	0.88
13q12.13	rs7989887	25681112	*PABPC3*	0.16185	0.087472	0.4964	6.93E-07	0.99
13q14.3	rs114375292	54413853	*LINC00558*	0.080925	0.030756	0.360387	7.25E-07	0.89
1q42.2	rs150148362	2.34E+08	*SLC35F3*	0.040462	0.008183	0.195651	8.16E-07	0.74
16p13.2	rs138031402	8960325	*CARHSP1*	0.043353	0.01044	0.232811	8.87E-07	0.77
7p13	rs146929064	44346336	*CAMK2B*	0.060694	0.020598	0.325486	1.32E-06	0.85
1q41	rs12058987	2.17E+08	*ESRRG*	0.083815	0.035553	0.402958	1.33E-06	0.98
8q24.21	rs7011138	1.28E+08	*PCAT1*	0.092486	0.19526	2.38087	1.35E-06	0.99
9p23	rs2039332	9138642	*PTPRD*	0.049133	0.011005	0.215339	1.48E-06	0.79
16p13.2	rs982855	8297685	*RBFOX1 /TMEM114*	0.572254	0.437641	0.581703	1.78E-06	0.99
6p21.1	rs6911069	43617687	*RSPH9*	0.069364	0.025113	0.34561	1.92E-06	0.99
20p13	rs56040592	1611918	*SIRPG*	0.104046	0.051072	0.463459	1.98E-06	0.98
3p24.1	rs142836448	29398714	*RBMS3*	0.063584	0.153217	2.66475	2.26E-06	0.95
2p23.3	rs77639126	25598423	*DTNB*	0.072254	0.023984	0.315525	2.35E-06	0.91
2p23.2	rs13028204	28853137	*PLB1*	0.086705	0.041479	0.455814	2.39E-06	0.98
3q22.1	rs114999403	1.32E+08	*CPNE4*	0.049133	0.014955	0.293814	2.67E-06	0.97
20p13	rs77402555	876841	*ANGPT4*	0.046243	0.012415	0.259286	2.69E-06	0.89
5p13.3	rs72753560	31626036	*PDZD2*	0.086705	0.178612	2.29048	2.75E-06	0.97
22q12.3	rs150551568	33136068	*SYN3*	0.046243	0.011851	0.247359	3.12E-06	0.95
14q24.2	rs117272176	72902205	*RGS6*	0.046243	0.014108	0.295149	3.23E-06	0.77
2p33.3	rs1207426	2.06E+08	*PARD3B*	0.16763	0.096501	0.530358	4.1E-06	0.99

15q25.1	rs8032417	78569930	*DNAJA4*	0.393064	0.275959	0.588522	4.82E-06	1.0
10p14	rs7084131	8399121	*GATA3*	0.011561	0.068002	6.23842	4.83E-06	0.99

Table XIV. Exonic and splicing variants of 23 GWAS suggestive genes in 138 FC cases and controls from WES

Gene	**Loci**	**Number of tested variants**	***P*-value (SKAT)**
ESRRG	1q41	2	0.7073712
SLC35F3	1q42.2	1	0.002189217
DTNB	2p23.3	2	0.002747022
PLB1	2p23.2	20	0.1122009
PARD3B	2p33.3	13	0.9205563
RBMS3	3p24.1	1	0.354907
CCDC80	**3q13.2**	**8**	**0.0005070405**
CPNE4	3q22.1	1	0.7073712
PROM1	4p15.32	12	0.05065522
PDZD2	5p13.3	19	0.7212919
RSPH9	6p21.1	2	0.8510591
CARD11	7p22.2	4	0.8172528
CAMK2B	7p13	3	0.09754801
PTPRD	9p23	7	0.2814774
PABPC3	**13q12.13**	**54**	**0.0001097577**
RGS6	14q24.2	1	0.7073712
DNAJA4	15q25.1	2	0.5637608
RBFOX1	16p13.2	5	1
CARHSP1	16p13.2	2	0.3250554
ANGPT4	20p13	4	0.09886071
SIRPG	20p13	5	0.7073707
HLCS	21q22.13	6	0.6115834
SYN3	22q12.3	3	0.5083075

Table XV. *CCDC80* exonic and splicing variants in FC IA cases and controls

Position	**Annotation**	**MAF (CEU)**	**SNP**	**MAC in case**	**MAC in control**
3:112324305	c.T2812C:p.Y938H	0.01	rs114697626	0	4
3:112328856	c.C2394G:p.L798L	0.002	rs61732242	0	2
3:112357336	c.C1417T:p.H473Y	0.009	rs116307644	4	1
3:112357537	c.A1216G:p.R406G	0		0	1
3:112357730	c.G1023T:p.L341F	0.01	rs115738438	5	2
3:112357878	c.C875A:p.A292E	0		0	1
3:112358168	c.G585C:p.E195D	0	rs370290734	0	1

3:112358171	c.T582A:p.G194G	0.01	rs116819563	5	2
3:112358658	c.G95A:p.S32N	0.001	rs75074453	0	1
3:112358761	UTR5 c.-9G>T	0		0	1

Table XVI. FC sample demographics and the clinical information of IA cases

Sample demographic of cases and controls		
	Cases	**Controls**
Number	257	1192
Mean age of recruit (SD)	53.9 (11.2)	54.4 (16.8)
% Male	31.5%	60.6%
Clinical information of cases (212)		
	With family history	**Sporadic**
With clinical information	147	65
Multiple lesions (%)	55 (37.4)	19 (29.2)
SAH (%)	51 (34.7)	25 (38.5)
Hypertension (%)	60 (40.8)	31 (47.7)
Drinking (%)	33 (22.4)	26 (40.0)
Smoking (%)	89 (60.5)	50 (76.9)
Hypercholesterolemia (%)	29 (19.7)	24 (36.9)

CHAPTER 4: DISCUSSION

4.1: POPULATION GENETICS: STUDY OF NUNAVIK INUIT

Genetic studies of complex diseases sometimes neglect to characterize the studied population first before launch into their genetic risk. This can be critical as some populations are admixed while others are much more homogeneous, and every population can be very different from the others. Until recently, most genetic studies of complex traits have been conducted using samples of European ancestry and this is now limiting our efforts to validate if the detected risk variants from the general human genome are functional variants or not in other distinct and isolated populations, and vice versa. Experimental designs can thus be biased – especially before the era of 1KGP. Consequently, when a non-European ancestry population is examined to identify the genetic factors underlying a complex disease, it is important to have a thoroughly characterized genomic profile of the targeted population, before attempting to establish if variants are associated with a particular trait or disease.

The work presented in the first part of this thesis was focused on uncovering the unique genetic landscape of an arctic indigenous population. Nunavik Inuit represent an important human population that has adapted to an extremely harsh environment since the Siberian Neolithic, possibly more than 5,000 years ago. Until very recently, Nunavik Inuit have been subjected to a high level of genetic isolation; and has made this population a powerful model to discover certain disease genes. Unfortunately, currently available commercial SNP genotyping arrays are designed to only suit major present-day populations, especially European descendants. We therefore resorted to use a WES approach – which is not commonly used in the field of population genetics. The approach has provided us with critical information regarding rare variants across genes under

selective sweeps, which would have been neglected if only a more conventional genotyping approach of common variants had been used.

Our first objective was to get a comprehension of the genetic structures of Nunavik Inuit. In comparison to similar Arctic indigenous populations, such as Northeastern Siberians[39] and Greenlandic Inuit[14], the genomes of Nunavik Inuit contain much higher levels of Inuit ancestry due to limited admixture with the Europeans, i.e., 87.6% of the Nunavik Inuit have >99% Inuit ancestry, by comparison, only 12% of Greenlandic Inuit have >95% Inuit ancestry and ~60% Inuit ancestry in overall Northeastern Siberians. Taking into account evidences such as the high LD of Inuit (**Figure 4**) and the higher mutation load reported for dispersed Out-of-Africa populations[14; 394], we assumed the genomic landscape of Inuit completely differs from the major worldwide populations and deemed that disease gene discovery approaches needed to be adjusted so that they would be best suited to the Inuit population.

Following an overview of the current knowledge regarding the demography of Nunavik Inuit, the first chapter presented analyses aimed at revealing genetic signatures that underlie their adaptation. Although Fumagalli *et al.* reported in the Greenlandic Inuit study that the *FADS* gene family is accounted to be the most significant signals as the result of positive selection[29], the fact that this family of genes is critical in dietary fat metabolism had been previously discovered as under adaptive evolution in many other populations[46; 271]. The Metabochip used in the Fumagalli *et al.* study was custom-designed to capture loci associated with metabolic traits; and again it was designed using variants that were essentially identified in individuals of European ancestry. Thus their findings failed to include critical information regarding the Inuit genome; one of the most important Inuit variant *CPT1A* p.P479L, which follows strong selective sweep, was noticeably not reported in their results.

Our study avoided such potential bias by making a repertoire of all coding variants across the Inuit genome. Though high-throughput sequencing generates variant call data with less accuracy compared to SNP genotyping array, it captures population specific variants that can be critical for the identification of loci under positive selection, as well as disease predisposition. In order to minimize the false positive call rate and bias caused by inconsistent wildtype/homozygous calls, we have developed a pipeline that: 1) calculates the high-confidence regions of our Inuit samples according to the initial set of variants' calls, and 2) only performs variant calling within the high-confidence regions of Inuit and 1KGP as reference data.

We used F_{ST}-based population branch statistic (PBS) to identify targets of recent and localized adaptation of Nunavik Inuit genome[236]. This method measures the variant frequency changes that are fitting with recent selection events since the Inuit's genome has started to diverge from the Han Chinese at approximately 23 kya BP[29]. We have also identified even more recent selection events that took place following the divergence of Nunavik Inuit and Northeastern Siberians, possibly at 4-8 kya BP according to the age of the *CPT1A* p.P479L variant, which was successfully replicated using PBS method.

The *CPT1A* p.P479L variant represents a very particular genetic signature of Inuit and Northeastern Siberians as these are the only two existing populations that commonly carry this loss of function variant. **Chapter 2.3** is largely dedicated to the study of *CPT1A* variants, as well as other variants in the genes encode carnitine acyltransferase (CPT) family (*CPT1A, CPT1B, CPT1C, CPT2* and *CART)*. Because of their importance in the fatty acid synthesis, Inuit may carry rare and damaging variants in these genes, such as *CPT1A* p.P479L variant, with a derived allele frequency of 95% among Nunavik Inuit while all homozygous in ancestral allele in non-Arctic indigenous populations. The *CPT1A* p.P479L variant can probably serve as an indicator of the

percentage of Inuit ancestry in an indigenous population. Because of their low-carbohydrate and high-fat diet of Arctic indigenous populations, these individuals may be in a permanent metabolic state of ketosis-like; the p.P479L variant is estimated to lead to a decreased activity of CPT1A and become a positive selective pressure to these people. Three other missense variants, *CPT1C* p.T265M, *CPT2* p.R477W and *CRAT* p.S78F were also found unique to Nunavik Inuit and were significantly more deleterious than the missense variants of those genes found in concurrent Asian populations. The F_{ST} score of these *CPT* genes also suggested significant population separations between Nunavik Inuit and Asians. This study was just a preliminary exploration of genes involved in Inuit metabolic genes related to diet, thus we had also presented an in-depth analysis in **Chapter 2.2**, where we looked for genome-wide signatures in Nunavik Inuit that would be linked to their adaptation to their diet and the arctic climate.

The genome-wide PBS analysis revealed 30 new genes (**Table I**) to be under selective sweep in Nunavik Inuit with *CPNE7* (PBS=1.74) and *ICAM5* (PBS=1.16) as the most interesting ones. The latter being the outcome of more recent selection events, as supported by high PBS score between Nunavik Inuit and Northeastern Siberians (PBS=1.04). The function of the *CPNE7* product is not very well established, while one group reported its association with odontoblast differentiation[395], others also pointed out its role in inflammatory response and oncogenesis[396]. The product of *ICAM5* is an intercellular adhesion molecule protein that is expressed in the vascular endothelium and it has been associated with hematological disorders[397]. It is noteworthy that both *CPNE7* and *ICAM5* are involved in vascular cell adhesion processes and this suggests that lipid metabolism genes are not the only ones involved in the adaptation of Inuit population.

To further the examination of the genes revealed by the PBS analyses under selective sweep, we used qRT-PCR to measure their levels of expression in both Nunavik Inuit and French-

Canadians, the latter representing a control population living in a similar geographical area. The French-Canadian colonization took place much later than Nunavik Inuit and the adaptation of this population has not yet been as fully established as the one of Inuit. Both *CPNE7* and *ICAM5* showed significant difference in expression levels between Inuit and French-Canadian LCL, similar trend also appeared in the previously-identified gene *FADS1*. The functional validation presented in our study, at our best knowledge, was the first of its kind to have used in any population-based genetic study, which has provided direct biological proof of the adaption signals. However, a major pitfall of this validation is that most of the coding variants do not affect gene expression levels. Moreover, several functionally interesting genes which also have high PBS scores, such as *CYP11B1* and *CYP11B2* family on chr8 and *IMPDH1* on chr7, failed to have detectable RNA expressions in LCL cells, which, unfortunately, were the only Inuit cells available to conduct our expression measures currently. Because LCL are not an ideal cellular model to explore the biological functions of these targeted genes, we cannot determine whether the genes that failed to show differential expressions between Nunavik Inuit and French-Canadians are involved in the adaptation of Inuit.

We further hypothesized that in an isolated founder population, genetic loci containing selection signals may also harbor genes predisposing this population to certain complex diseases, which may be related to the recent changes of their living environment. As a result, we discovered a splicing variant rs2289367 in *CCM2* (MAF=0.2) that is associated with the risk of IA. This variant is observed at a much higher frequency in Nunavik Inuit (MAF=0.82, PBS=0.82), thus suggesting that this approach could be a successful example as another way of identifying disease risk genes in small, isolated population. *CCM2* is a candidate for familial cerebral cavernous malformations (CCM), which is another cerebrovascular disorder that may have etiological

continuity with IA. Unlike IA, CCM mostly follows autosomal-dominant inheritance and only 5–15% of familial cases cannot be explained by the three known CCM genes (*CCM1-3*). The genetics of CCM largely relies on ethnic differences. A study discovered that mutations in *CCM2* accounted the majority of Japanese CCM cases[398], which is different from what was observed in Caucasian cases. It is not surprising that the rs2289367 variant in *CCM2*, which appears to be under selection or genetic drift, may be associated with another cerebrovascular disorder (IA) in another distinctive population. Additional correlations between genes under selection in Nunavik Inuit (revealed in iHS test) and the risk of IA may also be speculated. Among the other genes for which we observed an association, *SERPING1* presents a reduced expression in Nunavik Inuit LCL (**Appendix 1**). In past studies multiple members of the *SERPIN* family[179; 199] were reported to contribute to the development of IA.

Interestingly a scenario similar to the one described above was also observed in the second part of our study focused on the genetic susceptibility of IA in French-Canadians (**Chapter 3.2**); when we discovered that *RNF213*, a gene supposed to be also under genetic drift (**Figure 13**), appeared to be associated with IA in the French-Canadian population. *RNF213* was identified as a causative gene for Japanese moyamoya disease (MMD), which suggested these two cerebrovascular disorders share some genetic similarities. Furthermore, the population background seems to additionally contribute to the risk of MMD, as 95% and 79% of familial and sporadic MMD in Japanese are respectively explained by *RNF213* variants. Japanese is one of the East Asian populations for which a founder effect is observed. While Caucasian MMD cases were not found to be associated with *RNF213*[399]. Evidence from two of our studies suggest a new hypothetical approach for the study of cerebrovascular diseases, one that considers how different

functional disruptions (different variants) of the same gene may lead to the different phenotypes (outcomes) across the same body system, depending on the ancestry of the populations studied.

There are other examples suggesting genes (variants) under weak selection or genetic drift in Nunavik Inuit might also to be associated with highly prevalent disease in this population. As reported in one of our studies, a nonsense variant (rs61736969) in *TBC1D4*[400], which is well studied in the Inuit population because of its association with elevated postprandial glucose and T2D, we validated the findings first reported by Moltke *et al.*[30]. *TBC1D4* rs61736969 is another Inuit specific variant with MAF varies between Inuit populations (13.2% in Alaskan Inuit, 16.3% in Nunavik Inuit and 38.9% in un-admixed Greenlandic Inuit). Unlike the *CPT1A* p.P479L variant, the *TBC1D4* rs61736969 does not present in Northeastern Siberian populations, instead having a very low frequency in Native Americans (0.6%). We took a step further and looked into ancient American genome data[237], one of which traces back to an individual (6,260-5,890 cal BP, mtDNA haplogroup D4h3a7) from today's British Colombia. This individual has been described to have had a diet rich in marine protein, and also carried the alternative allele of *TBC1D4* rs61736969, indicating the variant may have occurred in the New World of Pleistocene America.

Several hypotheses can be proposed to explain how disease risk alleles happened to be located in the evolutionary adaptive signals. Williamson *et al.* mentioned that selective sweeps may "carry tightly linked, initially rare, deleterious, and potentially disease-causing variation to relatively high frequencies" (adaptive hitchhiking)[401]. Alternatively according to Dudley *et al.*, "high-frequency disease-associated variants can be the result of past adaptive events that now have detrimental consequences in our present contexts" (thrifty gene hypothesis and hygiene hypothesis)[402]. The functional relevance of the genes we observed to be within the Nunavik Inuit adaptation signals is likely due to the original adaptive events, especially in the cases of metabolic

adaptations, and this probably is one of the reasons behind the differences of CVD and IA prevalence among today's world populations. However, the thrifty gene hypothesis has its limitations, some of the signals we discovered using the PBS model thus may not only reflect the result of a positive selection, but instead have a neutral origin which has accumulated to high frequency through bottleneck effect (genetic drift). This scenario could have very well been at play for Nunavik Inuit and the recent changes of their lifestyle disrupted the neutrality of those variants and lead to their increased risk of developing CVD and IA. Nevertheless, because the adaptive variants have a higher probability of becoming pathogenic in altered environments (by comparison to previously neutral alleles), additional studies will be necessary to accurately interpret the enrichment of CVD and IA functional relevant genes in Nunavik Inuit selection signals.

Several weaknesses of this study still remain to be addressed. First, there may be some ascertainment bias due to the way of sampling – Inuit individuals were largely recruited as families and this may have led to an inflation when some of them are more related. However, the fact that any two individuals have at least one IBD>222 cM, even when they were recruited in different villages, indicates that such inflation is negligible in a population with strong founder effect. Nevertheless, the increased relativeness may have led to a certain bias in some analysis (e.g. ancestry estimation). Therefore, further tests and additional samples are needed to validate whether there is indeed an admixture of Native American ancestry in Nunavik Inuit as indicated in our study (**Figure 2**). While the increased relativeness of our cohort might not have influenced the adaptation signals we identified, additional numbers of Nunavik Inuit (especially in villages that we have not sampled from) could greatly strengthen the power to identify variants related to certain disease risk based on our PBS-based method. Also obtaining WES/whole genome sequencing

(WGS) data of other Canadian and Greenlandic Inuit could help in determine the population split time and to map a clearer migration routes.

The combination of SNP-chip genotyping and WES is probably the most efficient way by far to gather critical genetic information. However, some statistical tests (e.g. demographic inference and estimation of the selection starting time) perform better when using WGS data. A new human genome reference panel focused on Arctic indigenous population could also become available as more Inuit WGS is carried out, this would greatly enhance the imputation accuracy of genetic studies conducted for indigenous populations. Moreover, there may be other selective sweeps happening in the non-coding regulatory regions, which may be of equal importance towards the human evolution and adaptation as the coding variants, and we would not be able to detect these unbiasedly with the data currently available to us.

WGS is also ideal for the detection of copy number variation (CNV), and these have been shown to be associated with many complex diseases, and also to be subjected to natural selection during human evolution. A recent study describing the general spectrum of CNVs among global populations indicated that CNVs behave differently than SNV[403]. CNVs change more drastically over short period of time and consequently their differences across populations are much larger. Moreover, duplications and deletions also exhibit different functions in human genome adaptations as the number of duplications are much more than deletions, being the target for adaptive selection. Sudmant *et al*. also reported population stratification in a duplication in *CLPS*, a colipase functions in dietary metabolism of fatty acids[403], which happens to be one of the genes under selection in Nunavik Inuit discovered in our study (**Appendix 1**). The most interesting result of CNV studies is that they may explain part of the missing heritability of complex diseases, including cancer and

psychiatric diseases. Therefore, it would be complementary to our previous focus of SNV in order to explore the genetics of disease risks in Nunavik Inuit.

Another potential contribution to the population genetic study of Nunavik Inuit is examination of their mtDNA. Not only is it interesting to establish the mtDNA haplogroup of today's Nunavik Inuit, it is also important considering that in our study of Inuit adaptation to cold and diet. It has been observed that genes acting in energy generation pathways (mitochondrial coupling, ATP production and oxidative phosphorylation) have highly distinctive genetic profiles (**Appendix 1**). Mitochondria are structures within cells that convert the energy from food into ATP and it would not be surprising that a selection of mtDNA variants exist in Nunavik Inuit. Moreover, some of these variants could be deleterious and linked to some of the increased disease risks observed in this population. A recent study that examined the entire mtDNA genome of Greenlandic Inuit[404] revealed that Greenland Inuit are mostly descendent from three main haplogroups: A2a, A2b1 (16,200 years BP), and D4b1a2a1 (11,000 years BP). This confirmed the previous findings regarding sex biased Inuit-European admixture and it showed there was also lack of genetic continuity between Palo-Eskimos (D2a1) and Greenland Inuit. Our results which focused on autosomal DNA indicate that mtDNA findings may not likely to be the same in Nunavik Inuit, an examination of their mtDNA would thus complete the genetic picture of their adaptation. Some connections may exist between Nunavik mtDNA haplogroups, CNV and Inuit specific variants in the regulatory regions, and the genes we have thus far observed to be under selection signals. Overall it could be their collective contribution that affects the intricate energy generation mechanism as it has underlined the evolution and adaptations, and even the disease predispositions of Nunavik Inuit.

4.2: NEW RISK GENES FOR INTRACRANIAL ANEURYSM IN FRENCH-CANADIANS

IA is a very common but complex disorder that threatens the general population and more especially the elderly. It is estimated that as much as 10% of the population may harbor asymptomatic IA, depending on age, gender, as well as ethnicity[96]; though most of them remain unruptured and undetected. There are several risks factors and both the environment and genetic component are deemed to play a significant role. Nevertheless, it has been proven in the past that identifying the genetic risk factors of IA is a challenging task.

As IA is such a common disorder, GWAS seems to be the optimal approach – and several multi-ethnical GWAS have indeed discovered six major candidate regions including 9p21.3 (*CDKN2BAS*), 8q11.23-12.1 (*SOX17*)[207], 10q24.32 (*CNNM2*), 13q13.1 (*STARD13*), 18q11.2 (*RBBP8*)[320] and 4q31.22 (*EDNRA*)[210]. While the power of detection behind those loci was high enough for them to have been replicated in the follow-up studies, together they only explain 4.1-6.1% of the IA heritability[89].

It is possible that the genetic risk for IA differs across populations with different ancestries. It could be the consequence of an accumulation of low-frequency founder variants, some of which would contribute to an increased burden of variants in IA-susceptibility genes in a population predisposed to this condition. If such a scenario proves to be true it would explain the missing heritability of GWAS loci. In support of this a GWAS study that focused only on a cohort of Finnish IA case reported completely different risk loci (2q23.3, 5q31.3 and 6q24.2)[89]. However even with the increased power of imputation, GWAS, which are based on SNP genotyping platforms, are not an ideal approach to identify population specific variants, especially if those happen to be under a recent selective sweep in a high-risk population.

We decided to use WES in familial IA cases to capture population specific coding variants, even though usually high-throughput sequencing technologies are not best suited for the gene discovery of a complex disease, especially when heavily affected by genetic heterogeneity and incomplete penetrance. In order to minimize the false negative rate in risk variant discovery, we lowered the stringency in the variant prioritization step to allow the presence of low penetrance variants. We hypothesized that it would not be the increased frequency of common variants that would account the genetic heritability of IA in French-Canadians, thus our focus was on the rare variants specific to this population. We discovered that there seems to be a genetic continuity among cerebrovascular diseases as we identified French-Canadian IA variants in *RNF213*, a gene previously established to be a susceptibility factor for a rare type of cerebrovascular malformations, MMD, which was predominately observed in East Asians. Our study showed that variants located in different functional regions of RNF213 can in fact be associated with different types of cerebrovascular diseases in different populations as they affect the function of RNF213 in a distinct manner. Indeed we observed an increased activity in an ATPase assay where recombinant AAA+ domains of RNF213, carrying two variants (p.Arg2438Cys and p.Ala2826Thr) that were observed exclusively in French-Canadian IA case. By opposition, the Japanese founder mutation (p.Arg4810Lys) predisposing to MMD leads to a reduced ATPase activity[334]. Such divergent ATPase activities could be at the base of two distinct pathogenic mechanisms where the cerebrovascular walls would be affected differently in MMD and IA.

Our findings suggest that approximately 10% of French-Canadian familial IA cases might be explained by rare functional variants in *RNF213*, and that still leaves significant portion of the IA heritability to be discovered. At this point we only investigated *RNF213* in the French-Canadian cases and it remains to be established whether *RNF213* can be extended to IA cases from other

populations, something which at the very least should be done in population of European descents that are ethnically closest to French-Canadians. *RNF213* is a large gene, for which many private variants are listed in the NHLBI Grand Opportunity Exome Sequencing Project (GO-ESP) database. It is however important to consider that the ESP project is known heavily aggregated with cardiovascular and atherosclerosis patients, whom are supposed to have increased risk of IA as well. Moreover, the prevalence of unruptured IA can be 5-10% in general population, thus some private variants in GO-ESP database might be associated with increased IA risk as well. Further studies are necessary to determine whether *RNF213* is also an IA risk gene in non-Asian populations. This study, on the other hand, demonstrated that a WES approach can also be useful for studying common diseases under a right hypothesis.

To further our investigation of IA in the French-Canadians, we also performed a GWAS in an attempt to detect potential risk loci that were not discovered by WES. We used a large cohort (1,992) of control individuals to compensate the limited power of our relatively small IA cases (257) forming the discovery cohort. We identified a genome-wide significant loci at 3p14.2 ($p=4.66\times10^{-9}$) after imputation, which is located in the intronic region of an established tumor suppressor gene *FHIT*[374]. However, none of the loci previously reported in large and IA GWAS cohorts with multiple ethnicity had even reached the promising level of association (5×10^{-6}) observed in our study. Further analyses revealed another gene, *CCDC80,* which is located in a promising locus at 3q13.2 ($p=1.16\times10^{-7}$). *CCDC80* appears to present an increased mutation burden in IA patients when all the rare variants from the French-Canadian WES cohort are collapsed together (SKAT, p=0.0005), thus making it an additional interesting candidate.

Similar to the previous Finnish-only study[89], the replication rate of previous GWAS loci was also poor in our study. Only two SNPs, the well-established rs1333040 located in 9p21

(p=0.008565, no adjustment) and rs10958409 in 8q11 (p=0.006042, no adjust) showed significant association, however far from reaching genome-wide significant level. Apart from these two SNPs, none of the top SNP identified across other IA GWAS loci (4q31, 7p21, 10q24, 13q13 and 18q11), which were previously replicated at least twice, were found to show any significant association in our study. A region in 4q31 showed significant association ($p=4.8\times10^{-5}$), however this region is 440 kb upstream to the top SNP of *EDNRA* found in other studies[211].

The low replication rate of IA loci in our French-Canadian GWAS could be due to our limited number of cases. It is also possible that the previously conducted larger GWAS only revealed association signals, which are shared between populations, thus loci that would be more strongly associated with a higher IA risk in specific populations remain to be found, especially in populations that are not of Japanese or Finnish origin. The two largest IA GWAS were conducted in earlier days of the genome-wide research (2008[207] and 2010[320]). One may speculate if their results would have to be different, had population stratification been included as a covariate in the discovery stage of these studies. Nonetheless, previously established GWAS loci collectively explain only a very small portion of the IA heritability. Even in an independent follow-up Finnish study, only 2.5% of the heritability was explained, lower than the 6.1% indicated in the Finnish GWAS discovery study[89].

The criteria behind the selection of IA cases in our French-Canadian GWAS included a large number of cases with familial aggregation and this might also explain our identification of novel loci. An investigation of the genetic risk score for previous identified GWAS loci failed to discover significant associations when examining familial IA cases[213], suggesting additional potential risk loci of familial IA remain to be discovered.

Our GWAS identified two genes (*FHIT* and *CCDC80*) that may provide additional evidence to the genetic risk of familial IA in French-Canadians. *FHIT* has been associated with stress response and hemodynamic traits in populations from the Saguenay-Lac-St-Jean region, in which the residents are predominantly French-Canadian[376]. *CCDC80* is a cGMP signalling effector and affects the growth of vascular smooth muscle cells, and it is also associated with pulmonary arterial hypertension through its modulating effect on the expression of *endothelin-1* and *COL1A1,* two genes known to be critical in endothelial cells formation[405]. Low frequency SNPs were identified as top association signals in both *FHIT* and *CCDC80*, and these SNPs were identified through imputation based on the Haplotype Reference Consortium (HRC) dataset, which regroups 39 million SNPs to impute the haloptypes[390]. Low frequency variants have been shown to be informative in GWAS that were examining founder populations, as some of these were found representing high risk loci in Finnish IA cohort[89].

While it does not explain all heritability of IA in the French-Canadian population, *RNF213*, *FHIT* and *CCDC80* may collectively contributing to a fair share of the genetic etiology underlying IA in the French-Canadians. The studies of common diseases in founder populations offer an avenue that can be particularly useful and a number of simulations have shown that complex diseases are influenced by rare susceptibility alleles at many loci, as opposed to the GWAS-derived common disease/common variant (CD/CV) hypothesis. The CD/CV hypothesis cannot explain most of the complex diseases, as susceptibility alleles of complex diseases may have appeared more recently, and some of the more recent variants also came to be geographically restricted because of genetic drift and selective pressure. Therefore, the two studies we presented in the second part of this thesis highlight the importance of an approach to identify associations between low frequency variants specific to population and a complex disease, such as IA.

We also concluded from our studies that it may be beneficial to apply population stratification as covariates in large scale whole genome disease studies (e.g. GWAS). Such stratification should also be applied in higher resolution, which could be done using variant Site Frequency Spectrum (SFS) instead of PCA to avoid not only inflations but also potential false negatives brought by non-informative genotype data. New algorithms that introduce population specific data in the disease gene discovery may help to answer questions brought up by population heterogeneity and stratification.

Finally, towards the end of this thesis, we explored some new territories to relate the genetic susceptibility of IA and other complex traits with the course of human evolution. The risk of complex diseases owes to the human genetic diversity, particularly in metabolic (coronary artery disease, T2D, Stroke), immunological (TB, malaria, IBD) and neuropsychiatric phenotypes, which could be brought by either natural selection, genetic drift or admixture. As discussed above, past selective pressures might trigger local adaptations, which in turn might change the functional pathways and somehow lead to altered risk predisposition to certain diseases[406]. Admixture events might also be playing a role - a recent study discussed the presence of Neanderthal alleles in the contemporary human population and it appears that even small, archaic admixture can influence the risk to some diseases (immunology, metabolism, depression and addiction) in modern humans[407].

Natural selection signatures have been reported to overlap with many complex diseases (e.g. hypertension). Lipid metabolism is also one of the two key pathways that is subjected to human evolution, the other being inflammatory response[408]. Hypertension and inflammatory response have been known to affect the vascular wall remodelling of IA, and through evidence of atherosclerotic plaque formation in the affected vascular walls recent studies have also highlighted

the role of lipid metabolism in the pathogenesis of IA[256]. There are strong evidences of IA coexisting with many types of cardiovascular disorders, which therefore might also subject to human adaptive evolution. Thus, it is likely that we may also find new IA susceptible genes in the selection signatures of high risk populations, such as Japanese, Finnish, French-Canadians and Nunavik Inuit.

In the future, another important focus of IA genetic study is likely to concern the detection of somatic mutations. In one of our early IA experiment, we hypothesized that the pathogenesis of IA might be explained by the two-hit hypothesis. However, the analysis of this experiment was hindered by technical difficulties, the primary reason being that WES was unfit for detecting variants at very low mutation frequency, which we could not distinguish from sequencing artifacts at an average coverage of 100X. Some tissues affected by somatic mutations are easier to isolate from normal tissue (e.g. carcinoma cells), however the procedure is more difficult with aneurysmal lesions. Moreover, there are lack of potential candidate genes to be examined for the presence of somatic mutation, without which the current detection level of bioinformatic pipeline cannot accurately handle low frequency somatic mutations.

Considering that localized weakness is a key to aneurysmal formation, it is likely that somatic mutations could very well be involved in the pathogenesis of IA. A number of other cerebrovascular malformations are caused by the second somatic hit in their risk genes, for instance cerebral cavernous malformations (*CCM1-3*)[409], hemangioma and capillary malformations (*GNAQ*)[410], venous malformations (*PIK3CA*)[411] and arteriovenous malformations (*RASA1*)[412]. However, by opposition to those largely congenital cerebrovascular malformations, the development of IA can be affected by multiple factors and further investigations are warranted to establish whether or not IA is also associated with somatic mutations. Nevertheless, the discovery

of new candidate genes (e.g. *RNF213*) opens avenues for the testing of a somatic mechanism of the genes at play in IA.

Epigenetic modifications might also be another underlying cause leading to the development of IA. Studies have shown that in hypertension subjects DNA methylation affects genes involved in smooth muscle cell functions and in aortic aneurysm methylation affects genes encoding product involved in inflammatory response[413]. It was also reported that an enrichment of methylation profile variabilities is found across genomic signatures of natural selection and sometimes such methylation variabilities are linked to nearby genetic variants[414]. Thus it would be interesting to test if epigenetic profiles that are associated with adaptive selections are also at play for some of the increased predisposition to complex diseases like IA.

4.3: PERSPECTIVES: HUMAN EVOLUTION AND COMPLEX DISEASES

This thesis brought up three major hypothetical theories in addition to the findings described in each chapters. With the first hypothesis, we discovered that there seems to be certain genetic and phenotypic continuities in the disease phenotypes affecting the same system, in which the pathogenicity of variants within the same risk gene may vary according to the population background. Other genetic, environmental or epigenetic factors may also be the forces that trigger the direction of the mutational effect, thus leading to different outcomes of the diseases or traits. That is to say, if population stratifying is not applicable in the genetic study of a complex disease, then pooling together similar, overlapping phenotypes of multiple diseases may actually increase the power of detection in certain circumstances. Furthermore, mapping interactions involving risk genes, population genetic background and environmental factors would represent an important step to follow up the gene discovery and it could greatly impact the area of complex diseases studies. As more supporting evidence will be gathered regarding the ethnic background of patients and their susceptibility to a particular disease, it will provide important guideline for treatment, one can envision to achieve a better medical outcome.

The following hypothesis highlights the importance of integrating population genetic data into the statistical analysis of complex disease. Although currently there is a lack for algorithms that taking this into account, we have observed significant accumulation of population specific, low frequency variants using WES and SNP-chip datasets and some of these are contributing to the IA genetic etiology. Taking advantages of the fast progression of HTS technology and the exponentially increasing amount of human genome data, many groups are beginning to report new methods to help with the prioritization of mutation intolerant genes and to characterize

populations' variant landscapes. Undoubtedly, such algorithms will represent great additions to current bioinformatics pipeline as these will help complex disease studies by highlighting and incorporating population specific variants and genes in the progress of disease gene discovery.

The last hypothesis concerned past adaptive events that might be the roots of many complex diseases today. Selection on SNV, CNV or mtDNA represent driving forces that have affected the evolution of complex diseases and have led to the increased prevalence of some of these diseases in distinct populations by making its individuals more susceptible to them; although changes of CNV and mtDNA are thought to be much more drastic and assumed to be under less selective stress. Complex diseases to which humans are particularity susceptible (e.g. psychiatric disorders) have been suggested to also arise from certain ancient evolutionary events[415]. Past GWAS successfully testified the validity of the CD/CV hypothesis by discovering multiple risk loci based on large, multi-ancestry populations. But perhaps a new strategy could be developed in order to incorporate population genetic covariates (e.g. selection signals) in association studies, and this could fill the gaps left by GWAS and help to discover the missing heritability of some complex diseases. Just as admixture mapping has solved many problems in hunting for disease risk genes that were brought by the diverse ancestries in an admixed population, we believe a similar model that take into account the signals of natural selection could also yield significant results, especially in founder populations.

CONCLUSION

This thesis first described the genetic history and adaptive signatures of Nunavik Inuit, it also presented studies regarding the genetic susceptibility of IA in Nunavik Inuit and French-Canadian founder populations, which stressed the implication of their particular genetic backgrounds. Two major genes, *RNF213* and *CCM2* were discovered to be associated with IA in French-Canadians and Inuit, respectively, while both being risk genes for other cerebrovascular malformations in distinct populations. These new IA risk genes were revealed using population-based burden test that took into account the natural selection of population specific variants in the studied population, thus suggesting a new approach for the genetic investigation of complex diseases risk in isolated populations. Independently, additional discoveries were made in regard to the Nunavik Inuit population, including confirming the isolation and genetic homogeneity of this particular indigenous group and the identification of cardiovascular related genes located within the genetic signatures linked to their adaptations. This research further showcased the benefits from implementing (and sometime integrating) the datasets from HTS and high-density SNP array, this approach led to the identification of strong links between population specific variants and disease risks.

Appendix 1: Genetic Evidence for Cold and Diet Adaption in Canadian Inuit

Manuscript: Genetic Evidence for Cold and Diet Adaption in Canadian Inuit

Authors and affiliations:

Pingxing Xie[1,5], Sirui Zhou[2,5], Amirthagowri Ambalavanan[3,5], Amélie Quoibion[3,5], Patrick Dion[4,5], Guy A. Rouleau[3,4,5*]

[1] Faculty of Medicine, McGill University, Montreal, QC, Canada

[2] Department of Medicine, Université de Montréal, Montreal, QC, Canada

[3] Department of Human Genetics, McGill University, Montreal, QC, Canada

[4] Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada

[5] Montreal Neurological Institute and Hospital, Montreal, QC, Canada

Abstract

Nunavik Inuit have been challenged by the harsh conditions of the Arctic environment throughout their history. Due to the extremely cold weather, the traditional Inuit diet is unusually low in carbohydrates and high in animal protein and fat. Identifying the adaptive genetic factors associated with this population may provide insight into the health risks they are currently facing. In this study, we analyzed genome-wide SNP genotype data of 126 Inuit from the Nunavik region (northern Québec, Canada). Results from population structure analyses showed that Nunavik Inuit were genetically closest to Siberians living in the far northeastern Siberia, with shared genetic components from 42% to 65%. They also shared 25% genetic components with Na-Dene-speaking

Chipewyans residing in the northwestern and central Canada. Selective sweep analyses using iHS (integrated haplotype score) and XP-EHH (cross-population extended haplotype homozygosity) revealed that genes involved in protein metabolism (*SERPING1*), lipid metabolism (*APOBEC1* and *CLPS*) and body temperature regulation (*ADRA2A* and *ESRRG*) were located in the strongest selected regions. Compared to French Canadians, who live south to the Nunavik Inuit, the expression level of *SERPING1* in cells from Inuit was significantly reduced. Collectively, our study identified the genetic relationship of the Inuit and Siberians, and the genetic factors underlying the adaptation of the Inuit to their environment.

Introduction

Recent studies have found that a minimum of three gene flow events have happened from Siberia to the Americas through Bering Strait [1]. The first event happened at least 15,000 years ago. After they arrived, the majority travelled to the south. The second and third events happened about 4,500 and 1,000 years ago. They represent the settlement of the Native Americans in the North American Arctic. Paleo-Eskimos, who migrated to America during the second event, are genetically close to Na-Dene groups currently residing in the northwestern and central Canada[1; 25]. Paleo-Eskimos are not the direct ancestors of the modern Canadian Inuit. Modern Canadian Inuit are likely to be the descendants of the Thule people from Neo-Eskimo lineage, who came during the third migration event and replaced Paleo-Eskimos 700 years ago[9].

Currently, the Inuit represent approximately 89% of the inhabitants of the Nunavik region. Throughout their history, Nunavik Inuit have been challenged by the harsh conditions of the Arctic environment. Based on the data from Environment Canada (1981 to 2010)[416], in January and

February, the average daily mean temperature of Kuujjuaq, the administrative centre of Nunavik, was -25 °C, with wind chill commonly below -40 °C at night. As the extremely cold weather makes it impossible to cultivate plants, Nunavik Inuit are skilled fishermen and hunters of caribou and large marine mammals including seals and walrus. The traditional Inuit diet is unusually low in carbohydrates and high in animal protein and fat[417]. Nonetheless, several studies conducted in Greenland and Canada have reported that the age-adjusted mortality due to cardiovascular diseases of the Inuit was lower than that of the non-Inuit populations[418; 419], although the results were not replicated by other studies[420; 421]. Recently, the lifestyle of the Inuit has substantially changed with a decrease of physical activity and an increased consumption of Western food, which results in a dramatic increase in the prevalence of obesity, type 2 diabetes, and cardiovascular diseases[2].

How the Inuit genetically adapt to the cold environment and high protein and lipid diet remains largely unknown. A nonsynonymous variant in *CPT1A* (rs80356779) was previously suggested to be the target of a strong selective sweep in the Northeast Siberian populations[40]. *CPT1A* encodes a key enzyme in the carnitine-dependent transport across the mitochondrial inner membrane during long-chain fatty acids metabolism. We have also observed that the derived allele frequency of rs80356779 was very high in Nunavik Inuit (96%)[49]. It is possible that this allele is strongly selected in the Arctic populations, in part because of its association with increased serum high-density lipoprotein and thus protective effect of cardiovascular diseases[304; 305]. Recently, Fumagalli *et al.* identified a cluster of three fatty acid desaturase genes (*FADS1*, *FADS2* and *FADS3*) to be strongly selected in Greenlandic Inuit[29]. The *FADS* gene cluster may have effects on metabolic phenotypes, as well as height and weight.

In this study, we analyzed genome-wide SNP genotype data of 126 Inuit from Nunavik, Québec. For the 112 Inuit without evidence of mixed ethnicity, we performed population structure analyses and compared them with Native American and Siberian populations from Reich *et al*.[1]. Selection analyses identified that genes involved in protein metabolism (*SERPING1*), lipid metabolism (*APOBEC1* and *CLPS*) and body temperature regulation (*ADRA2A* and *ESRRG*) were located in the strongest selected regions. While the expression of these genes is regulated by physiological conditions and cell types, we found that the expression of *SERPING1* in cells from the Inuit was half of what is observed in the French Canadian population.

Results

Population structure analyses

In this study, we genotyped DNA samples of 126 Nunavik Inuit using Illumina HumanOmniExpress-12v1 platform containing 730,525 SNPs. After quality control steps, 112 Inuit without evidence of mixed ethnicity were left for further analyses. We first performed principle component analysis (PCA) of the 112 Nunavik Inuit, and Native Americans and Siberians without evidence of recent admixture of other ethnicities from the study of Reich *et al*.[1]. Unrelated individuals of CEU (Utah Residents with Northern and Western European Ancestry) and CHB (Han Chinese in Beijing, China) from HapMap phase 3 were used as reference populations. The first principal component (PC1) explained 33% of the variance in the data, separating Native American (including Nunavik Inuit) and Siberian populations from CEU and CHB. PC2 explained 20% of the variance, separating Nunavik Inuit from the other Native American and Siberian populations. Nunavik Inuit were genetically closest to Siberians, and they were quite distinct from the other Native American populations (**Fig. 1A**).

Admixture analysis was used to further explore the genetic relationships between Nunavik Inuit and indigenous peoples of the Americas and Siberia. At K=5 (**Fig. 1B**), Nunavik Inuit shares 65% genetic components with Eskimo-Aleut-speaking Naukans, as well as 42% genetic components with Chukchi-Kamchatkan-speaking Chukchi and Koryaks. All three populations are indigenous inhabitants of the far northeastern Siberia. We also observed that Nunavik Inuit share 25% genetic components with the Na-Dene-speaking Chipewyans residing in the northwestern and central Canada. Nunavik Inuit shared very few (< 8%) genetic components with the other Native American populations.

Selection analysis

To identify genomic regions under positive selection, we performed iHS (integrated haplotype score) and XP-EHH (cross-population extended haplotype homozygosity) analyses of genome-wide SNP genotype data of 39 unrelated Nunavik Inuit. A total of 12,649 and 12,773 windows, each with 200 kb region, were in the iHS and XP-EHH analysis, respectively. iHS analysis is most powerful to detect selective sweeps with moderate frequency, while XP-EHH has maximum power to detect close to fixed selective sweeps[422].

Gene ontology (GO) enrichment analysis of the 208 genes located in the top 1% iHS windows revealed significant enrichment of GO term "cellular senescence" (P=0.0076). Six of the 208 genes are known to encode proteins regulating cell cycle and/or apoptosis under cellular stress (*OPA1*, *SRF*, *MAPK14*, *ID2*, *ARNTL*, and *SMC6*). For the 268 genes located in the top 1% XP-EHH windows, GO enrichment analysis revealed significant enrichment of GO terms related to sensory perception of smell due to overrepresentation of olfactory receptor genes (P=9.2×10^{-6}). Olfactory receptor gene family contains more than 400 genes with a large number of genetic

polymorphisms[423; 424]. The significant GO term may reflect high genetic diversity and positive selection of certain olfactory receptor genes in Nunavik Inuit.

To focus on the most strongly selected genes, we further examined the genes located in the top 10 ranking 200-kb-windows in the iHS and XP-EHH analyses (**Table 1**). Forty-nine genes mapped to these windows. We extensively explored the functions of these genes in the published literatures. Because of the unique living environment and diet habit of Nunavik Inuit, we focused on genes involved in lipid and protein metabolisms, as well as body temperature regulation.

Two genes involved in lipid metabolism were strongly selected. *APOBEC1* was very close to (2.0 kb downstream) the 1st ranking iHS window chr12:7.6-7.8 Mb, and it was fully covered by the positively selected haplotype. The protein product of *APOBEC1* is a catalytic subunit of an apolipoprotein B (apoB) mRNA editing enzyme complex. In human small intestine cells, apoB mRNA editing enzyme converts mRNA of apoB-100 into apoB-48 by introducing a stop codon[425]. Intestinal apoB-48 is crucial in chylomicron assembly during fat absorption[426]. The second gene involved in lipid metabolism was *CLPS*. Although it was 34.9 kb upstream of the 8th ranking iHS window (chr6:35.8-36.0 Mb), *CLPS* was completely covered by the positively selected long-range haplotype in this region. *CLPS* is a small gene encoding pancreatic colipase preproprotein in exocrine pancreatic cells. In small intestine, trypsin splits colipase preproprotein into enterostatin and colipase. Enterostatin reduces the serum cholesterol level in mice on a high cholesterol diet[427]. Colipase is a cofactor of pancreatic lipase. It increases the efficiency of lipid hydrolysis by stabilizing the active conformation of lipase[428].

One gene involved in protein metabolism, *SERPING1*, was located in the 7th ranking iHS window (chr11:57.2-57.4 Mb). *SERPING1* encodes C1 serine protease inhibitor which inactivates

C1r, C1s, kallikrein, and factor XII in the complement system and blood coagulation cascade[429]. It also inhibits chymotrypsin[430], which catalyzes the hydrolysis of proteins in human small intestine.

Two genes regulating body temperature, namely *ADRA2A* and *ESRRG* were located in the 4th ranking XP-EHH window (chr10:112.8-113.0 Mb) and the 9th ranking iHS window (chr1:216.8-217.0 Mb), respectively. *ADRA2A* encodes the α2-adrenergic receptor, which regulates brown adipose tissue thermogenesis[431]. Its agonists, such as clonidine and xylazine, lead to dose-dependent hypothermia[432]. In human subcutaneous white adipose tissue, sympathetic nervous system activation inhibits lipolysis through α2-adrenergic receptor signalling pathway to maintain fat pads[433; 434], which is important in cold exposure[435]. *ESRRG* encodes the orphan nuclear receptor estrogen-related receptor γ (ERRγ). Overexpression of ERRγ stimulates nonshivering thermogenesis in brown adipocytes through upregulating uncoupling protein 1 (UCP1) and oxidative genes expression levels[436; 437].

Gene expression analyses of genes mapping in the top 10 iHS and XP-EHH windows

We investigated the expression levels of 29 of the 49 genes located in the top 10 iHS and XP-EHH windows, as well as *APOBEC1* and *CLPS* in 14 unrelated Nunavik Inuit and 15 unrelated French Canadians using RT-qPCR of EBV-transformed lymphoblastoid cell lines. In windows with more than one gene, those covered by the strongest selection signals were tested. Eighteen of the 31 genes were found to yield high quality RT-qPCR results (**Table 2**). Most genes encoding products involved in metabolism and body temperature regulation, including *APOBEC1*, *ADRA2A*, and *ESRRG*, are not expressed in lymphocytes[438], and thus their expression levels could not be evaluated. Compared to French Canadians, *SERPING1* was found to have significantly reduced

expression in Inuit. The expression ratio (Nunavik Inuit / French Canadians) of *SERPING1* was 0.57 (*P*=0.05, **Fig. 2**).

Identification of coding variants in the top 10 iHS and XP-EHH windows

Among the 112 Inuit without evidence of mixed ethnicity or poor DNA quality, we performed exome sequencing in 104 of them. To identify functional variants that were likely to be the targets of selection sweeps, we searched for variants located in the exons and the splicing sites (≤ 6 bp away from an exon) of the 49 genes mapping to the top 10 iHS and XP-EHH windows in 104 Nunavik Inuit. We also searched for variants in *APOBEC1* and *CLPS*. A total of 39 qualified variants were identified in the top 10 iHS windows, and another 45 variants in the top 10 XP-EHH windows. Three of the 84 variants had allele frequencies ≥ 30% in Nunavik Inuit, and ≤ 10% in ExAC database (**Table 3**). However, only one of them was nonsynonymous (rs41265137). This variant is located in *SCN3A* which encodes a subunit of voltage-gated sodium channels. The minor allele frequency of rs41265137 was 42% in Inuit and 1.9% in ExAC database ($P < 0.0001$, Chi-square test).

Strong selection sweeps in chromosome 11q

Recently, Cardona *et al*. investigated genome-wide SNP genotype data of northeast Siberian populations and observed positive selection signals in chr11:66-69 Mb, which contains 79 protein-coding genes[39]. Further analysis of the sequence data of this region suggested that the nonsynonymous variant rs80356779 located in *CPT1A* may be the target for this strong selective sweep[40]. We have previously reported that the derived allele frequency of rs80356779 was 96% in our Inuit subjects[49], which is higher than that observed in northeast Siberian population (68%)[40]. However, in this study, our analyses based on extended haplotype homozygosity failed to identify

any positive selections in or around *CPT1A*, which is located in window chr11:68.4-68.6 Mb. In XP-EHH analysis, we observed selective signals in chr11:67.4-67.8 Mb, which ranked 16th genome-wide. This same region was the 1st ranking window in XP-EHH analysis in northeast Siberian populations in the study of Cardona *et al*.[39]. In our study, the genes with the strongest selection signal in chr11:67.4-67.8 Mb were *ALDH3B1* and *ALDH3B2*, both of which encode aldehyde dehydrogenase. Aldehyde dehydrogenase protects the cells from the lipid peroxidation-derived aldehydes including trans-2-octenal, 4-Hydroxynonenal, hexanal, H_2O_2 and menadione[439]. In addition, we found two nonsynonymous variants in *ALDH3B2* that were enriched in the Inuit (chr11:67431974_C>T and chr11:67432821_C>T). They had extremely low minor allele frequencies in ExAC database (0.9% and 0.008%), while their frequencies in the Inuit were both 33% ($P < 0.0001$ for both variants, Chi-square test).

Discussion

In this study, we investigated genome-wide SNP genotype data of 112 Nunavik Inuit without evidence of mixed ethnicity. Population structure analyses showed that Nunavik Inuit shared 42-65% genetic components with the indigenous peoples living in the far northeastern Siberia. They also shared 25% genetic components with the Na-Dene-speaking Chipewyans residing in the northwestern and central Canada, which may reflect a genetic admixture between Paleo-Eskimos/Na-Dene and Neo-Eskimos. Nunavik Inuit shared very few genetic components (< 8%) with the other Native American populations living in the United States and South America, which was consistent with previous studies[1; 9; 25].

Selection tests in Nunavik Inuit identified that genes involved in protein metabolism (*SERPING1*), lipid metabolism (*APOBEC1* and *CLPS*) and body temperature regulation (*ADRA2A*

and *ESRRG*) were located in the top 10 most selected 200-kb-windows in the iHS and XP-EHH analyses. C1 serine protease inhibitor encoded by *SERPING1* inhibits chymotrypsin. The reduced level of *SERPING1* observed in Inuit cell lines, if it also occurs in the duodenum, would allow more activated chymotrypsin for protein hydrolysis during a high-protein traditional Inuit diet. In terms of fat digestion, both genes have well established roles in lipid metabolism. The product of *APOBEC1* truncates apoB-100 to form apoB-48 in small intestine cells. ApoB-48 assembles chylomicrons, which transport fat from the intestine to liver during fat absorption. *Apobec1*$^{-/-}$ mice had defective transport of triacylglycerol from intestinal mucosa to lymph after a high-fat diet[440]. *CLPS* encodes two proteins, enterostatin and colipase. Enterostatin induces early satiety to prevent overconsumption of fat[441] and reduces serum cholesterol level[427]. Colipase increases the efficiency of lipid hydrolysis. Compared to wild type mice, *Clps*$^{-/-}$ mice had deceased body fat, impaired fat absorption, elevated cholesterol and reduced triglycerides[442]. The two genes regulating thermogenesis, *ADRA2A* and *ESRRG*, play important roles in preventing hypothermia. Both of them have been shown to regulate nonshivering thermogenesis in brown adipocytes[431; 436; 437]. In human subcutaneous adipocytes, when sympathetic nervous system is activated by cold exposure, subcutaneous fat pads can be maintained through α2- adrenergic receptor signalling pathway[433-435]. Increased nonshivering thermogenesis and preservation of the subcutaneous fat would give the Inuit survival advantages to live in the artic environment.

Gene expression analysis of EBV-transformed Inuit and French Canadians lymphoblastoid cell lines found that the expression of *SERPING1* was reduced by almost 50% in Nunavik Inuit. For *CLPS*, based on the results from the genotype-tissue expression (GTEx) project, the expression level of *CLPS* was 2.5×10^4 times higher in pancreas than in EBV-transformed lymphoblastoid cell line[438]. Therefore, although the expression patterns in cell lines were similar between Inuit and

French Canadians, a small change in gene expression in cell lines might reflect a large difference in the exocrine pancreatic cells. Only by examining pancreas tissues would we be able to confidently compare the expression pattern of *CLPS* between the Inuit and non-Inuit populations. According to the GTEx Consortium database[438], *APOBEC1* is only expressed in small intestine (terminal ileum), *ADRA2A* is largely expressed in adipose tissues, and *ESRRG* is expressed in multiple tissues including brain and heart. However, none of them are expressed in EBV-transformed lymphoblastoid cell lines, so we were unable to measure their expression levels in the Inuit.

Recently, Fumagalli *et al*.[29] studied Greenlandic Inuit and identified strong positive selection signals in a cluster of three fatty acid desaturase genes (*FADS1*, *FADS2* and *FADS3*) located in chr11:61.4-61.6 Mb. We did not observe any selective signals in or around this region in Nunavik Inuit. It is important to point out that the methodology of our study and Fumagalli *et al*.[29] were different. We used unbiased genome-wide SNP genotype data (730,525 SNPs) and selection analyses based on extended haplotype homozygosity. In Fumagalli *et al*.[29], genotype data was generated using Illumina MetaboChip, which contains 196,725 SNPs from GWAS studies associated with atherosclerotic and cardiovascular disease traits. In addition, their selective sweep results were from population branch statistics (PBS), a test based on SNP frequency differences rather than haplotype homozygosity. Therefore, our results are not comparable with those from Fumagalli *et al*.[29].

We searched for functional variants that were likely to be the targets of selection sweeps in the top 10 iHS and XP-EHH windows. By examining the exons and the splicing sites of genes located in the windows, we only observed one nonsynonymous variant (rs41265137) in *SCN3A*

that had allele frequencies ≥ 30% in Nunavik Inuit and ≤ 10% in ExAC database. Due to the moderate allele frequency, we could not conclude that rs41265137 was the target of the selective sweep in this region. Because this study only investigated SNP genotype data, an important direction for future work will be to analyze exome sequencing or even genome sequencing data to identify variants that are strongly selected in the Inuit.

Materials and Methods

Subjects

We recruited 126 Inuit from thirteen villages of Nunavik, Québec, Canada. All participants were self-reported as Inuit. Written informed consent was obtained from all participants. The study was approved by University Ethics Committee and the Nunavik Nutrition and Health Committee.

SNP genotyping and quality control

DNA was extracted from whole blood samples and genotyped on two 96-well plates of Illumina HumanOmniExpress-12v1 platform containing 730,525 SNPs at the McGill University and Génome Québec Innovation Centre. SNP genotypes were called using the genotyping module in the GenomeStudio software (Illumina). Quality control steps were performed using PLINK[293]. We filtered out SNPs a) located on non-autosomes, b) having call rate < 96%, c) having plate bias $P < 10^{-5}$, c) lacking ancestral allele information in the dbSNP database. In the end, a total of 697,804 SNPs were kept for analysis.

For the quality control steps, we first used PLINK to identify and remove five samples with ambiguous sex. To identify Inuit of mixed ethnicity, we merged genotype data of Nunavik Inuit and unrelated individuals of CEU, JPT-CHB and YRI from HapMap phase 3. SNPs with one allele

of A and the other allele of T, and SNPs with one allele of C and the other allele of G were removed before merging, because it was impossible to know whether the two sets of SNPs were concordant in terms of positive or negative strand, as they were genotyped on different chips. Then we pruned a set of 50,848 unrelated SNPs with $r^2 < 0.1$ within a 50-SNP sliding window, and performed multidimensional scaling (MDS) analysis using PLINK[293]. Ten individuals (one was also identified to have ambiguous sex) had mixed ethnicity of Caucasians and Inuit. ADMIXTURE analysis[231] using the same data confirmed the results. After deleting DNA samples with ambiguous sex and mixed ethnicity, 112 Nunavik Inuit were left, all of whom had < 4.5% mixture from other ethnicity (0.7% on average). All samples had call rate > 99%.

Next we calculated genome-wide alleles shared identical-by-descent (IBD) and PI_HAT by PLINK to exclude Nunavik Inuit who were closely related. We randomly removed one member of a pair of Inuit if the PI_HAT value was greater than 0.18, which is between the second (PI_HAT=0.25) and the third degree relative (PI_HAT=0.125). After the above quality control steps, 39 unrelated Inuit were left for selection analyses.

Population structure analyses

We requested access to the genome-wide genotype data of 52 Native American and 17 Siberian groups from the study of Reich *et al*.[1]. Two hundred and ten Native Americans and Siberians with no evidence of recent admixture with other ethnicities were included in the population structure analyses with the data of 112 Nunavik Inuit who had no evidence of mixed ethnicity. Due to the limited number of individuals in each Native American and Siberian population from Reich *et al*.[1], we grouped them into nine linguistic families based on the original study using Greenberg's classification[1], namely, Andean (n=6), Central-Amerind (n=21), Chibchan-Paezan (n=47),

Chukchi-Kamchatkan (n=31), Equatorial-Tucanoan (n=66), Eskimo-Aleut (n=19), Ge-Pano-Carib (n=7), Na-Dene (n=2), and Northern-Amerind (n=11). Among the 50,848 unrelated SNPs from the above MDS analysis, 32,204 of them were also genotyped by Reich *et al*.[1] which were extracted for principal component analysis (PCA). PCA was conducted using the Eigensoft package[274; 329] to identify population structure of Nunavik Inuit, Native Americans, and Siberian populations. Unrelated individuals of CEU (n=106) and CHB (n=80) from HapMap phase 3 were used as reference populations. The same set of SNPs were also analyzed using ADMIXTURE[231] to estimate genome-wide admixture pattern of Nunavik Inuit compared to the nine linguistic families of Native Americans and Siberians.

Selective sweep analyses

We phased the 697,804 SNPs of the 39 unrelated Inuit using Beagle 4.0[275], where phased 1000 Genomes Project phase 3 data was used as the reference panel. The ancestral state of each SNP was retrieved from dbSNP human Build 142. The iHS (integrated haplotype score)[443] was calculated using an R package rehh[444]. iHS scores were normalized based on allele frequencies using the same R package. The iHS test result was ranked by the fraction of SNPs in each 200 kb window with the absolute value of iHS > 2. After excluding windows with less than five SNPs, a total of 12,649 windows were kept in the results of the iHS analysis.

For the XP-EHH (cross-population extended haplotype homozygosity) analysis[445], we used 85 unrelated CHB individuals from 1000 Genomes Project phase 3 as the reference population. XP-EHH analysis was done using the same method described in Pickrell *et al.*[422] (http://hgdp.uchicago.edu/Software/). XP-EHH scores were normalized with mean of 0 and

standard deviation of 1. A total of 12,773 windows each with 200 kb region were in the analysis. Results were ranked by the highest normalized XP-EHH score in each 200 kb region.

Genes located in top 1% iHS and XP-EHH windows were retrieved based on consensus coding sequence (CCDS) release 19. Gene ontology (GO) enrichment analysis was performed using the tool provided by Gene Ontology Consortium (http://geneontology.org/).

Gene expression analysis by real-time quantitative reverse transcription-PCR (RT-qPCR)

We tested the expression levels of 31 genes located in the top 10 iHS and XP-EHH windows in EBV-transformed lymphoblastoid cell lines of 14 unrelated Nunavik Inuit and 15 unrelated French Canadians. In windows with more than one genes, those with the highest selection signals were tested. cDNA was generated from RNA extractions prepared from lymphoblastoid cell lines using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). RT-qPCR was performed using the TaqMan Gene Expression Assay on QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems). Three independent RNA extractions and RT-qPCR experiments were done for all 29 lymphoblastoid cell lines. In every RT-qPCR experiment, the same cDNA sample was tested in triplicate.

Next, gene expression levels were normalized to the RNA polymerase II polypeptide A gene (*POLR2A*). For each gene, ΔCt values of 14 Nunavik Inuit were compared with those of 15 French Canadians by Bayesian analysis with Markov Chain Monte Carlo (MCMC) using R package rjags (http://www-ice.iarc.fr/~martyn/software/jags/). MCMC simulation allows the data and its variability to generate sampling distributions of the maximum likelihood estimator without strong prior or test assumptions. *P* values were calculated from 100,000 comparisons of this

estimator between Nunavik Inuit and French Canadians. Expression ratio of each gene was calculated using the Livak method ($\Delta\Delta$Ct)[276].

Exome sequencing and identification of selected variants in the top 10 iHS and XP-EHH windows

Among the 112 Inuit who had no evidence of mixed ethnicity or poor DNA quality, we performed exome sequencing for 104 of them. The detailed methods have been described before[49]. In short, all DNA samples were captured by Agilent SureSelect 50 Mb (V4) chip and sequenced on the Illumina Hiseq 2000. Three samples were barcoded to be sequenced in one lane to ensure an average coverage depth of 100 fold. Raw sequence reads were aligned to GRCh37 using Burrow-Wheeler Aligner[288] and variants were called by Genome Analysis Toolkit using standard methods[267]. ANNOVAR[289] was used to perform variant annotation.

We searched for variants located in the exons and the splicing sites (≤ 6 bp away from an exon) of genes covered by the top 10 iHS and XP-EHH windows. Quality filters were set at sequencing depth ≥ 20x and variant frequency ≥ 25%, with genotype quality ≥ 10. Allele frequencies of the variants observed in the 104 Inuit subjects were compared with those found in ~60,000 individuals of Exome Aggregation Consortium (ExAC, release 0.2). To identify coding variants that were positively selected in the Inuit, we focused on variants with allele frequencies ≥ 30% in Nunavik Inuit subjects, and ≤ 10% in ExAC database. All qualified variants were manual inspected using the Integrative Genomics Viewer (IGV)[292].

Acknowledgements

We thank all the Nunavik Inuit individuals and their families participating in this work. The study was approved by McGill University Institutional review board # A06-M31-97.

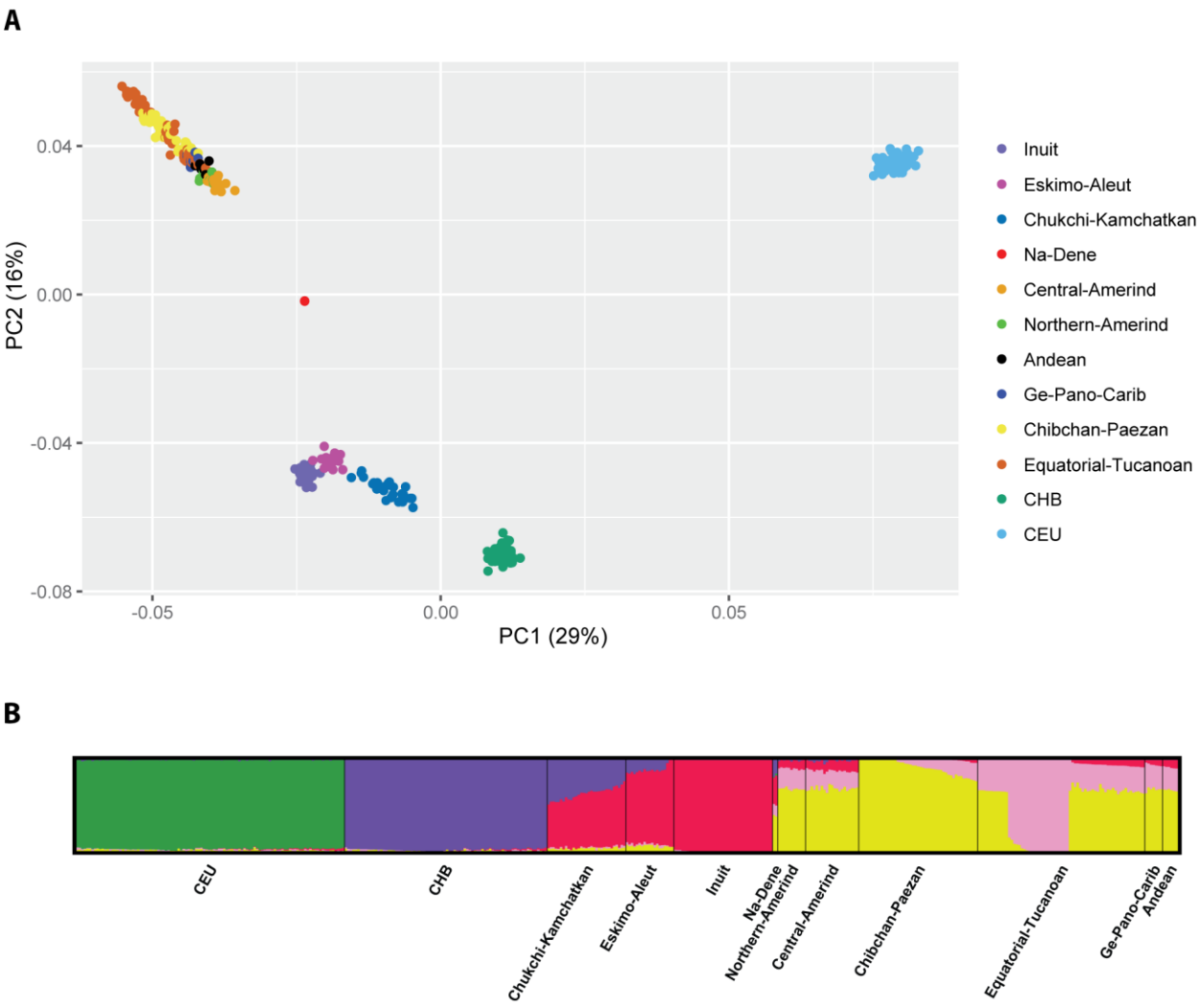

Fig. 1: Principal component analysis (**A**) and Admixture analysis (K=5, **B**) of 112 Nunavik Inuit who had no evidence of admixture, and 210 Native Americans and Siberians with no evidence of recent admixture with other ethnicities from the study of Reich *et al*. Native Americans and Siberians were grouped into nine linguistic families based on Reich *et al*. using Greenberg's classification. (Andean: Aymara, Quechua and Yaghan; Central-Amerind: Pima, Tepehuano and Zapotec; CEU: Utah residents with Northern and Western European ancestry; CHB: Han Chinese in Beijing; Chibchan-Paezan: Bribri, Cabecar, Embera, Guaymi, Maleku, Teribe and Waunana; Chukchi-Kamchatkan: Chukchi and Koryak; Equatorial-Tucanoan: Chane, Guahibo, Guarani, Jamamadi, Karitiana, Palikur, Parakana, Piapoco, Surui, Ticuna and Wayuu; Eskimo-Aleut:

Naukan; Ge-Pano-Carib: Arara, Toba and Wichi; Na-Dene: Chipewyan; Northern-Amerind: Kaqchikel, Maya and Mixe.)

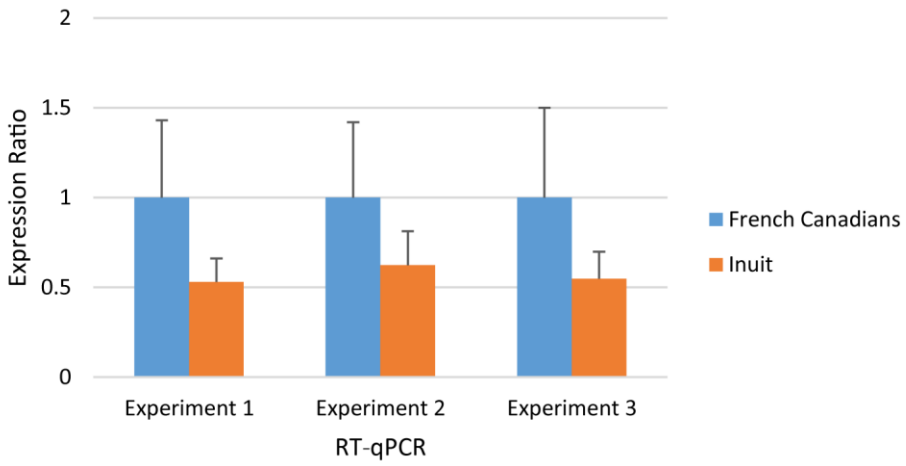

Fig. 2: RT-qPCR experiments identified that the expression ratio (Nunavik Inuit / French Canadians) of *SERPING1* was 0.57 (P=0.05). Three independent RNA extractions of EBV-transformed lymphoblastoid cell lines (14 unrelated Nunavik Inuit and 15 unrelated French Canadians) and RT-qPCR experiments were performed.

Table 1: Top 10 iHS and XP-EHH windows in Nunavik Inuit, and protein-coding genes that map in the respective window.

Rank	Window (hg 19), Mb	Genes
iHS		
1	chr2:24.4-24.6	*FAM228A, ITSN2*
1	chr2:72.4-72.6	*EXOC6B*
1	chr4:47.8-48.0	*CORIN, NFXL1, CNGA1*
1	chr5:42.0-42.2	
1	chr10:94.6-94.8	*EXOC6*
1	chr12:7.6-7.8	*CD163*
7	chr11:57.2-57.4	*RTN4RL2, SLC43A1, TIMM10, UBE2L6, SERPING1*
8	chr6:35.8-36.0	*SRPK1, SLC26A8, MAPK14*
9	chr1:216.8-217.0	*ESRRG*
10	chr14:60.8-61.0	*C14orf39, SIX6*
XP-EHH		
1	chr12:56.2-56.4	*ORMDL2, DNAJC14, MMP19, WIBG, DGKA, PMEL, CDK2, RAB5B, SUOX*
2	chr12:56.0-56.2	*OR10P1, METTL7B, ITGA7, BLOC1S1, RDH5, CD63, GDF11, SARNP*
3	chr4:18.6-18.8	
4	chr10:112.8-113	*ADRA2A*
5	chr13:53.2-53.4	*HNRNPA1L2, SUGT1, LECT1*
6	chr2:166.0-166.2	*SCN3A, SCN2A*
7	chr15:58.0-58.2	*GCOM1, POLR2M*
8	chr20:21.0-21.2	
9	chr11:14.2-14.4	*SPON1, RRAS2*
10	chr19:54.2-54.4	*NLRP12, MYADM, PRKCG*

Table 2: RT-qPCR results of 29 out of 49 genes mapping in the top 10 iHS and XP-EHH windows, as well as *APOBEC1* and *CLPS*. In windows with more than one genes, those with the highest selection signals were tested. Three independent RNA extractions and RT-qPCR experiments were done for EBV-transformed lymphoblastoid cell lines of 14 unrelated Nunavik Inuit and 15 unrelated French Canadians. Eighteen of the 31 genes had high quality RT-qPCR results. Expression ratio (RQ) with $P < 0.05$ were highlighted in bold.

Gene	Position (hg19)	Rank of genomic window	RQ_1	RQ_2	RQ_3	RQ mean
FAM228A	chr2:24397972-24414567	iHS_1	NA	NA	NA	NA
ITSN2	chr2:24425735-24583397	iHS_1	0.84	0.8	0.95	0.86
EXOC6B	chr2:72406444-73053177	iHS_1	0.8	0.67	0.83	0.77
CORIN	chr4:47596018-47840059	iHS_1	NA	NA	NA	NA
NFXL1	chr4:47849258-47916633	iHS_1	0.98	1.02	0.8	0.93
EXOC6	chr10:94608225-94819251	iHS_1	0.8	0.92	0.87	0.86
CD163	chr12:7623412-7656414	iHS_1	NA	NA	NA	NA
APOBEC1	chr12:7801996-7818502	iHS_1	NA	NA	NA	NA
RTN4RL2	chr11:57228339-57244384	iHS_7	NA	NA	NA	NA
SLC43A1	chr11:57252004-57282358	iHS_7	1.18	0.95	1.21	1.11
TIMM10	chr11:57295936-57298232	iHS_7	1.17	0.95	1.02	1.05
UBE2L6	chr11:57319128-57335180	iHS_7	0.96	0.78	1.11	0.95
SERPING1	chr11:57365027-57382326	iHS_7	**0.53**	**0.62**	**0.55**	**0.57**
CLPS	chr6:35762759-35765121	iHS_8	0.88	0.66	0.74	0.76
SRPK1	chr6:35800811-35888957	iHS_8	1.1	0.89	1.11	1.03
SLC26A8	chr6:35911291-35992305	iHS_8	NA	NA	NA	NA
MAPK14	chr6:35995454-36079013	iHS_8	1.05	0.77	1.04	0.95
ESRRG	chr1:216676588-216896814	iHS_9	NA	NA	NA	NA
C14orf39	chr14:60902674-60952764	iHS_10	NA	NA	NA	NA
SIX6	chr14:60975938-60978525	iHS_10	NA	NA	NA	NA
ORMDL2	chr12:56211806-56214959	XP-EHH_1	1.55	0.95	1.1	1.2
DNAJC14	chr12:56214744-56223420	XP-EHH_1	0.97	0.87	0.94	0.93
MMP19	chr12:56229214-56236767	XP-EHH_1	0.92	0.94	0.71	0.86
SARNP	chr12:56146247-56211540	XP-EHH_2	1.07	0.93	1.13	1.04

ADRA2A	chr10:112836790-112840662	XP-EHH_4	NA	NA	NA	NA
LECT1	chr13:53277400-53313947	XP-EHH_5	NA	NA	NA	NA
SCN3A	chr2:165944030-166060577	XP-EHH_6	1.59	1.5	1.44	1.51
POLR2M	chr15:57998901-58001556	XP-EHH_7	1.14	0.85	0.99	0.99
SPON1	chr11:13984184-14289679	XP-EHH_9	NA	NA	NA	NA
RRAS2	chr11:14299466-14380730	XP-EHH_9	1.11	0.65	0.83	0.86
NLRP12	chr19:54296855-54327657	XP-EHH_10	NA	NA	NA	NA

Table 3: Coding variants of 49 genes located in the top 10 iHS and XP-EHH windows, as well as those in *APOBEC1* and *CLPS* were investigated in 104 Nunavik Inuit using exome sequencing method. Three variants were found to have allele frequencies ≥ 30% in Nunavik Inuit, and ≤ 10% in Exome Aggregation Consortium (ExAC, release 0.2) database.

Position (hg19)	**Variant**	**dbSNP**	**Gene**	**Variant Function**	**Inuit Allele Frequency**	**ExAC Frequency**
chr11:57268808	G>T	rs150660444	*SLC43A1*	intronic_splicing	0.67	0.012
chr2:165947079	C>A	rs41265137	*SCN3A*	nonsynonymous	0.42	0.019
chr19:54403925	T>C	rs2242244	*PRKCG*	synonymous	0.38	0

Appendix 2: Supplemental material for Chapter 2.2

Supplemental Figures

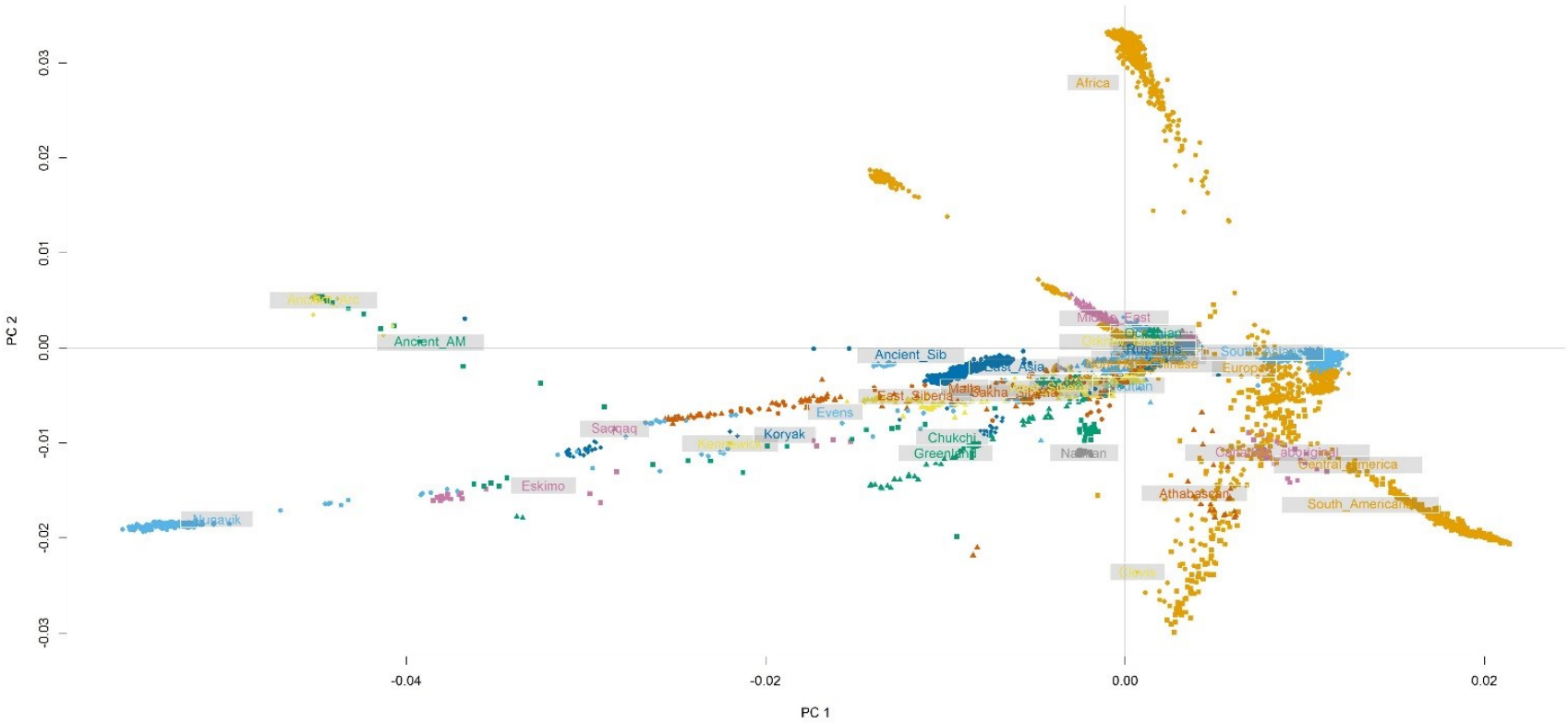

Figure S1: PCA result showing all populations included in the structural analysis.

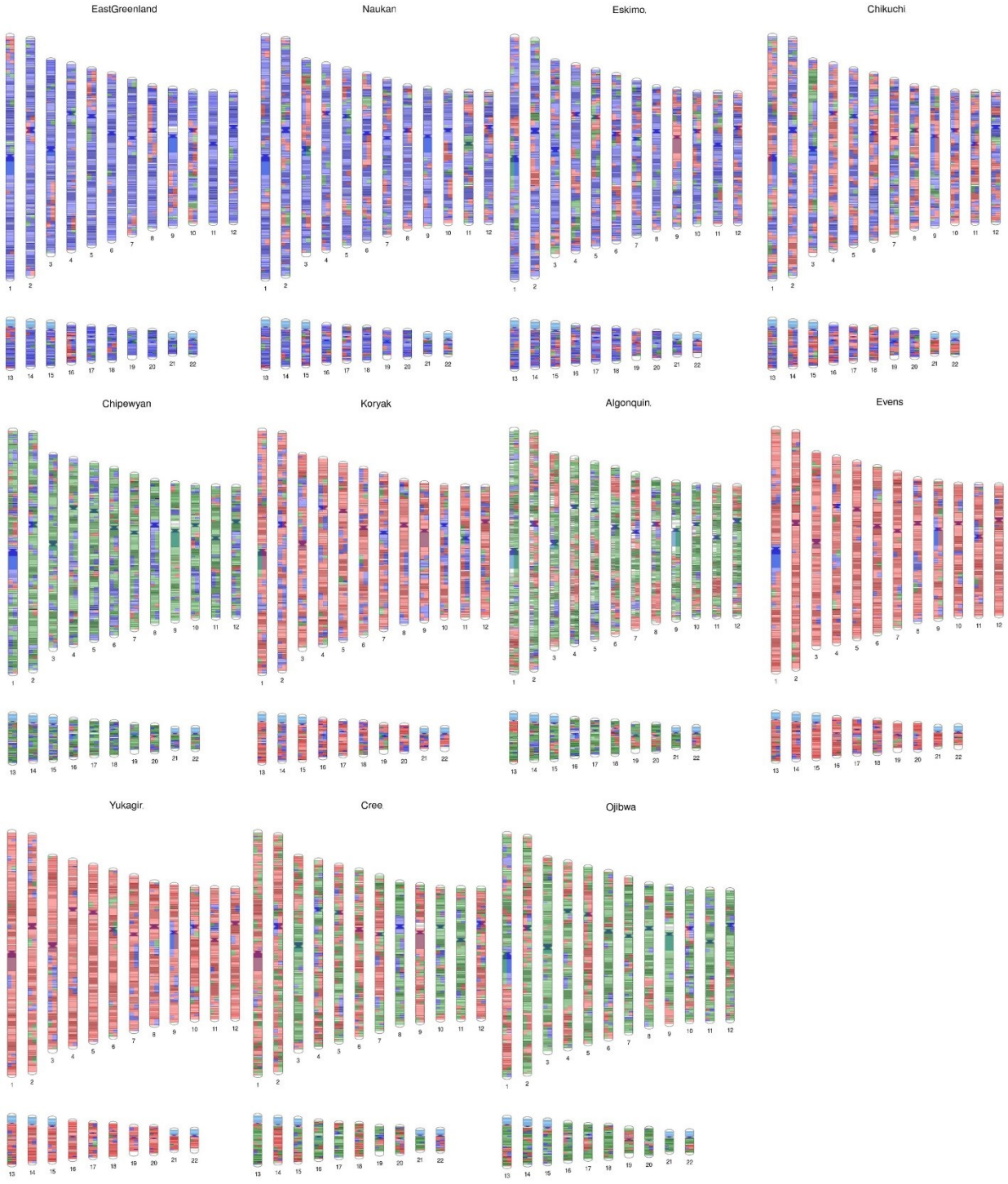

Figure S2: Genome-wide local ancestry of Arctic indigenous populations (blue: Inuit ancestry; red: European ancestry; green: Native American ancestry).

Figure S3: Nunavik Villages and sample recruitment (figure adapted from http://www.inuitfirstcanadians.com/)

Supplementary tables

Table S1: Study populations and controls

	Array based (no.populations)	**WGS / WES (no.populations) [Ancient name]**
Nunavik Inuit (current study)	170 (10)	114 (10)
Siberians (Cardona, 2014; Clemente, 2014)	218 (10)	25 (3)
Native Americans / Siberians (Reich, 2012)	2351 (66)	NA
Sakha-Siberians (Fedorova, 2013)	40 (9)	NA
Caucasus (Yunusbayev, 2012)	204 (13)	NA
Turkic (Yunusbayev, 2015)	322 (32)	NA
Siberia-New World (Rasmussen, 2010)	188 (14)	1 [Saqqaq]
Altai regions (Raghavan, 2014)	85 (9)	1 [Mal'ta]
Siberians (Raghavan, 2015)	20 (5)	9 current + 23 ancient
1000 Genome phase III	NA	2504 (26)
Ancient America (Rasmussen, 2014)	NA	1 [Clovis]
Ancient New World Arctic (Raghavan, 2014)	NA	34 (5)
Ancient Altai (Allentoft, 2015)	NA	5
Ancient America (Rasmussen, 2015)	NA	1 [Kennewick]
TOTAL	5422 (197)	

Table S2: 132 variants from PBS analysis met the empirical threshold of 99.9th.

chrom	pos	rs	MAF.Inuit	MAF.CEU	MAF.CHB	PBS	gene	function
11	68548130	rs80356779	0.956731	0.000505	0.000485	3.111911	*CPT1A*	missense
15	48426484	rs1426654	0.000481	1	0.029126	2.062288	*SLC24A5*	missense
3	12574298	rs735640	0.830097	0.000505	0.00049	1.756087	*TSEN2*	3-prime-UTR
16	89661807	rs12445560	0.947115	0.09596	0.126214	1.744347	*CPNE7*	synonymous
7	1.28E+08	rs72624969	0.975962	0.070707	0.286408	1.650973	*IMPDH1*	intron
7	1.28E+08	rs4731447	0.975962	0.076531	0.276699	1.648679	*IMPDH1*	intron
16	89662837	rs12447806	0.985149	0.111111	0.252427	1.634635	*CPNE7*	intron
16	89656162	rs139901937	0.802885	0.000505	0.000485	1.606432	*CPNE7*	intron
7	1.28E+08	rs2288553	0.966346	0.075758	0.281553	1.564838	*IMPDH1*	intron
12	56743044	rs2066815	0.793269	0.000505	0.009709	1.525998	*STAT2*	missense-near-splice
20	13550127	rs11697393	0.830097	0.020202	0.048544	1.524275	*TASP1*	intron
8	1.44E+08	rs5280	0.961538	0.015152	0.441748	1.521063	*none*	upstream-gene
16	89971316	rs140032244	0.802885	0.000505	0.024272	1.519637	*TCF25*	intron
8	1.44E+08	rs4534	0.961538	0.015152	0.446602	1.511838	*CYP11B1*	missense
16	90130139	rs4264393	0.805825	0.000602	0.007143	1.504595	*PRDM7*	missense
20	13514603	rs6033751	0.831731	0.020202	0.058252	1.499884	*TASP1*	intron
20	13463855	rs2273479	0.831731	0.015152	0.063107	1.498412	*TASP1*	intron
22	42276742	rs2228314	0.995192	0.252525	0.194175	1.48112	*SREBF2*	missense
19	55418054	rs2278428	0.956731	0.065657	0.320388	1.45223	*NCR1*	missense
15	78191608	rs58169519	0.870192	0.030303	0.135922	1.451833	*none*	intergenic
4	47578971	rs16851681	0.980769	0.232323	0.194175	1.433606	*ATP10D*	missense
16	89760474	rs12595968	0.947115	0.065657	0.31068	1.405871	*CDK10*	intron
16	89702620	rs2070992	0.947115	0.107143	0.257282	1.393311	*DPEP1*	intron
19	55417731	rs2278427	0.956731	0.080808	0.325243	1.391171	*NCR1*	intron
3	12858557	rs180768267	0.788462	0.000556	0.021739	1.389524	*CAND2*	missense
16	88909095	rs35137494	0.879808	0.156566	0.058252	1.379805	*GALNS*	intron
4	47548431	rs12498466	0.960784	0.206522	0.195	1.377129	*ATP10D*	intron
4	48566843	rs79352370	0.911458	0.084337	0.209677	1.377057	*FRYL*	intron
2	1.1E+08	rs3827760	0.004808	1	0.067961	1.348744	*EDAR*	missense
3	12641425	rs5746223	0.826733	0.09596	0.033981	1.344314	*RAF1*	intron
16	89294595	rs74466939	0.908654	0.050505	0.252427	1.338114	*ZNF778*	synonymous
16	89663072	rs659974	0.004808	0.858586	0.592233	1.315908	*CPNE7*	3-prime-UTR
13	53049267	rs34494025	0.740385	0.010101	0.000485	1.30311	*CKAP2*	synonymous

2	1.34E+08	rs55724759	0.028846	0.903409	0.570423	1.292132	*NCKAP5*	intron
13	53286822	rs62637607	0.735577	0.010101	0.000485	1.28517	*LECT1*	intron
1	1.59E+08	rs2276405	0.75	0.000505	0.029126	1.279471	*AIM2*	missense
13	53238148	rs112815564	0.73301	0.010101	0.000485	1.278989	*SUGT1*	intron
3	1.26E+08	rs4679121	0.826923	0.074713	0.067164	1.274227	*ZXDC*	intron-near-splice
8	1.44E+08	rs6432	0.95098	0.015152	0.519608	1.26684	*CYP11B2*	intron
22	40803112	rs55844816	0.879808	0.09596	0.169903	1.25915	*SGSM3*	intron
12	1.31E+08	rs1212936	0.966346	0.328283	0.160194	1.240326	*GPR133*	intron
8	1.44E+08	rs3764795	0.038835	0.954545	0.504854	1.237918	*GML*	missense
1	2.13E+08	rs3738800	0.024038	0.838384	0.631068	1.2193	*NSL1*	intron
16	89757132	rs12598665	0.947115	0.141414	0.305825	1.219236	*CDK10*	intron
16	89986154	rs885479	0.014423	0.921348	0.436782	1.199063	*MC1R*	missense
4	48610961	rs776604	0.081731	0.823171	0.76	1.196354	*FRYL*	intron
2	2.41E+08	rs77816205	0.754808	0.066327	0.000485	1.194793	*NDUFA10*	intron
13	52598189	rs17480245	0.705882	0.010101	0.000485	1.187379	*ALG11*	missense
13	76445155	rs9544057	0.052885	0.78	0.648649	1.185765	*none*	upstream-gene
5	54398339	rs3136555	0.798077	0.06701	0.082524	1.177476	*none*	upstream-gene
22	40800518	rs760701	0.879808	0.136364	0.169903	1.1761	*SGSM3*	intron
22	40800544	rs2235318	0.879808	0.136364	0.169903	1.1761	*SGSM3*	intron
6	49755030	rs28372930	0.932692	0.116162	0.320388	1.173198	*PGK2*	5-prime-UTR
13	52678741	rs34326189	0.711538	0.020202	0.000485	1.170403	*NEK5*	intron
11	61580635	rs174556	0.004808	0.689655	0.664706	1.169148	*FADS1*	intron
2	1.34E+08	rs76742066	0.028846	0.89375	0.5	1.166751	*NCKAP5*	intron
19	10402938	rs1056538	0.019231	0.616162	0.81068	1.163544	*ICAM5*	missense
19	10403368	rs2228615	0.019231	0.616162	0.81068	1.163544	*ICAM5*	missense
13	52513266	rs7334118	0.706731	0.005495	0.000588	1.152985	*ATP7B*	missense
16	27246617	rs17856580	0.009615	0.773684	0.604167	1.147446	*NSMCE1*	missense
19	10370542	rs11115	0.004808	0.656566	0.718447	1.147055	*MRPL4*	3-prime-UTR
19	10403947	rs2569702	0.019231	0.611111	0.805825	1.146108	*ICAM5*	intron
15	44198288	rs76332091	0.68932	0.000505	0.004902	1.143801	*FRMD5*	intron
11	61624414	rs174602	0.932692	0.191919	0.257282	1.140539	*FADS2*	intron
1	2.13E+08	rs2291772	0.024038	0.838384	0.57767	1.128437	*FLVCR1*	intron
1	2.13E+08	rs17019870	0.024038	0.838384	0.57767	1.128437	*FLVCR1*	intron
1	2.13E+08	rs3768548	0.033654	0.858586	0.57767	1.117401	*FLVCR1*	intron
19	55420924	rs3765014	0.043269	0.836735	0.627451	1.110749	*NCR1*	intron
8	1.44E+08	rs2294008	0.033654	0.594118	0.786885	1.105699	*PSCA*	5-prime-UTR
1	1.52E+08	rs4845606	0.802885	0.085859	0.106796	1.105564	*RORC*	intron
1	2.13E+08	rs17019873	0.029126	0.838384	0.57767	1.098868	*FLVCR1*	intron
16	89212200	rs4302032	0.9375	0.136364	0.354369	1.096629	*ACSF3*	intron

11	61569830	rs174546	0.004808	0.639535	0.648649	1.096416	*FADS1*	3-prime-UTR
13	52715168	rs55969405	0.706731	0.005814	0.000833	1.096289	*NEK3*	intron
19	746206	rs2384776	0.932692	0.252525	0.223301	1.094707	*PALM*	intron
19	16630224	rs12460141	0.692308	0.005495	0.005952	1.093478	*CHERP*	intron
11	57268808	rs150660444	0.668269	0.000505	0.000485	1.088858	*SLC43A1*	intron-near-splice
4	47560386	rs4298115	0.019231	0.570707	0.81068	1.088358	*ATP10D*	intron
22	40801312	rs12483888	0.879808	0.106061	0.247573	1.086952	*SGSM3*	intron
13	52544577	rs151202111	0.695	0.010989	0.000704	1.085456	*ATP7B*	intron
11	61551927	rs174536	0.009615	0.676768	0.674757	1.085322	*MYRF*	intron
1	2.13E+08	rs41300993	0.04902	0.861446	0.589109	1.079095	*FLVCR1*	intron
7	66281040	rs1984651	0.884615	0.121212	0.247573	1.078964	*GTF2IRD1P1*	intron
1	2.13E+08	rs2291771	0.033654	0.833333	0.582524	1.076281	*FLVCR1*	intron
11	47311905	rs147735512	0.663462	0.000505	0.000485	1.074561	*MADD*	intron
6	7574536	rs7741957	0.854369	0.000505	0.315534	1.074353	*DSP*	intron
1	2.13E+08	rs1890932	0.033981	0.838384	0.57767	1.073821	*FLVCR1*	intron
8	1.44E+08	rs2978982	0.033654	0.593023	0.769841	1.071862	*PSCA*	synonymous
4	47561179	rs11730590	0.019231	0.575758	0.796117	1.071372	*ATP10D*	intron
11	61570783	rs174547	0.004808	0.639175	0.673684	1.069415	*FADS1*	intron
16	49823239	rs2271063	0.859223	0.107143	0.212766	1.066081	*ZNF423*	intron
11	57080882	rs145672238	0.663462	0.000505	0.004854	1.063441	*TNKS1BP1*	missense
15	1.01E+08	rs76411473	0.826923	0.036082	0.218447	1.06048	*ADAMTS17*	intron
1	2.04E+08	rs3737975	0.052885	0.942105	0.485294	1.060047	*LAX1*	intron
13	42335222	rs9566844	0.941748	0.328283	0.194175	1.056225	*VWA8*	intron
5	6605271	rs506416	0.014851	0.70202	0.650485	1.054262	*NSUN2*	intron
5	54319997	rs3822315	0.043269	0.811828	0.613861	1.051998	*none*	upstream-gene
17	47123824	rs4265867	0.663462	0.005102	0.004854	1.051564	*IGF2BP1*	intron
11	61560081	rs174538	0.004808	0.661616	0.650485	1.051003	*TMEM258*	5-prime-UTR
19	16639121	rs12459056	0.745192	0.076923	0.025974	1.050362	*CHERP*	intron
12	52369300	rs2242106	0.932692	0.20202	0.300971	1.050174	*ACVR1B*	intron
19	44351694	rs2195980	0.038462	0.813131	0.601942	1.048763	*ZNF283*	missense
2	1.15E+08	rs13418557	0.870192	0.050505	0.305825	1.047977	*ACTR3*	intron
4	47561141	rs13106043	0.019231	0.565657	0.791262	1.04723	*ATP10D*	intron
11	47304564	rs143653424	0.653846	0.000505	0.000485	1.04657	*MADD*	intron
3	12628920	rs2290159	0.821053	0.222222	0.038835	1.040277	*RAF1*	intron
12	96180692	rs59294014	0.908654	0.015152	0.470874	1.040168	*NTN4*	intron
3	12861600	rs181307051	0.663462	0.000538	0.000568	1.03994	*CAND2*	missense
12	53344318	rs2070875	0.067308	0.994949	0.412621	1.036721	*KRT18*	intron
3	12633083	rs2290160	0.822115	0.227273	0.033981	1.0341	*RAF1*	intron
7	44612229	rs2289057	0.841346	0.020202	0.281553	1.032202	*DDX56*	synonymous
1	2.13E+08	rs10864004	0.024038	0.806122	0.553398	1.031811	*C1orf227*	missense

11	61557826	rs102274	0.004808	0.637755	0.656863	1.02932	*TMEM258*	intron
11	68678962	rs560096	0.048077	0.883838	0.538835	1.028293	*IGHMBP2*	missense
13	28239940	rs11029	0.942308	0.353535	0.18932	1.02751	*POLR1D*	synonymous
11	61547237	rs108499	0.004808	0.641414	0.650485	1.022169	*MYRF*	intron
12	1.31E+08	rs1195923	0.028846	0.571429	0.796117	1.018131	*GPR133*	intron
6	7580958	rs2076299	0.855769	0.035354	0.305825	1.01754	*DSP*	intron
19	16666347	rs73514949	0.692308	0.010204	0.048544	1.017043	*SLC35E1*	intron
11	61549025	rs174533	0.004808	0.636364	0.650485	1.015209	*MYRF*	intron
19	16617486	rs74418457	0.692308	0.005051	0.053398	1.013774	*C19orf44*	intron
5	6620459	rs2288446	0.033654	0.712121	0.665049	1.012473	*NSUN2*	intron
7	44609752	rs2304373	0.836538	0.020202	0.281553	1.012413	*DDX56*	intron
7	44611009	rs3793250	0.836538	0.020202	0.281553	1.012413	*DDX56*	intron
12	8808222	rs79153086	0.745192	0.055556	0.087379	1.011511	*MFAP5*	intron
11	61551356	rs174535	0.004808	0.636364	0.645631	1.008394	*MYRF*	synonymous
3	12618269	rs2442805	0.826923	0.232323	0.053398	1.007626	*MKRN2*	intron
5	6604970	rs6888601	0.019231	0.680412	0.643564	1.006427	*NSUN2*	intron
5	76734260	rs33205	0.759615	0.050505	0.116505	1.005718	*WDR41*	intron
5	6602544	rs2303711	0.033654	0.712121	0.660194	1.005363	*NSUN2*	intron
3	1.08E+08	rs3957558	0.868932	0.109756	0.28	1.000842	*MYH15*	intron
19	10395683	rs5498	0.019231	0.585859	0.737864	0.997435	*ICAM1*	missense

Table S3: Variants under different selection strength between Nunavik Inuit with Europeans and Nunavik Inuit with Northeast Siberians.

chrom	pos	rs	PBS (NUI-CHB-CEU)	gene	function	AF (GI)	PBS (GI)	PBS (NUI-NES-CHB)	MAF.NES	CADD
1	1.59E+08	rs2276405	1.279471376	*AIM2*	missense	0.8611	2.536098	0.634696	0.28	16.34
3	12858557	rs180768267	1.389523908	*CAND2*	missense	#N/A	#N/A	1.094545	0.12	16.91
3	12861600	rs181307051	1.039940373	*CAND2*	missense	0.3889	1.067382	0.625435	0.14	25.7
4	47578971	rs16851681	1.433606237	*ATP10D*	missense	#N/A	#N/A	1.073339	0.46	0.09
8	1.44E+08	rs3764795	1.237917802	*GML*	missense	#N/A	#N/A	0.368187	0.28	15.17
8	1.44E+08	rs4534	1.511837689	*CYP11B1*	missense	#N/A	#N/A	0.495403	0.32	0.009
11	61551356	rs174535	1.008393664	*MYRF*	synonymous	1	0.76399	0.284129	0.08	0.924
11	68548130	rs80356779	3.111911226	*CPT1A*	missense	#N/A	#N/A	0.839809	0.32	18.21
11	68678962	rs560096	1.028292975	*IGHMBP2*	missense	0	0.954138	0.15293	0.12	1.24
12	56743044	rs2066815	1.5259979	*STAT2*	missense-near-splice	#N/A	#N/A	0.839179	0.24	25
13	52513266	rs7334118	1.152985299	*ATP7B*	missense	0.4722	1.25654	0.613102	0.2	11.53
13	52598189	rs17480245	1.187378938	*ALG11*	missense	0.4722	1.212806	0.532059	0.22	0.017
13	52715168	rs55969405	1.096288839	*NEK3*	intron	0.4722	1.212806	0.571077	0.22	#N/A
13	53049267	rs34494025	1.303110102	*CKAP2*	synonymous	#N/A	#N/A	0.502239	0.3	4.588
16	89661807	rs12445560	1.744346679	*CPNE7*	synonymous	#N/A	#N/A	0.48916	0.74	9.53
16	89986154	rs885479	1.199063038	*MC1R*	missense	0.8889	0.57	0.375297	0.18	9.495
16	90130139	rs4264393	1.504595068	*PRDM7*	missense	#N/A	#N/A	0.039348	0.62	8.187
19	10395683	rs5498	0.997435183	*ICAM1*	missense	0.8056	0.424054	0.881544	0.4	0.754
19	10402938	rs1056538	1.163543685	*ICAM5*	missense	#N/A	#N/A	1.038068	0.44	16.96
19	10403368	rs2228615	1.163543685	*ICAM5*	missense	0.8056	0.5465	1.038068	0.44	12.92
19	55418054	rs2278428	1.452229608	*NCR1*	missense	#N/A	#N/A	0.510609	0.3	5.939

Table S4: Result of GO analysis of 1,596 genes with one or more coding variants with weak signals of selection (PBS>0.3)

GO biological process (PANTHER Overrepresentation Test)	Total (20814)	observed	expected	fold Enrichment	P-value
cell adhesion (GO:0007155)	1039	124	78.87	1.57	6.52E-03
biological adhesion (GO:0022610)	1044	124	79.25	1.56	8.25E-03
localization (GO:0051179)	4838	452	367.25	1.23	3.79E-03
cellular response to stimulus (GO:0051716)	6061	560	460.09	1.22	2.60E-04
signaling (GO:0023052)	5052	461	383.5	1.2	3.78E-02
single organism signaling (GO:0044700)	5049	460	383.27	1.2	4.61E-02
response to stimulus (GO:0050896)	7482	677	567.96	1.19	7.46E-05
single-organism cellular process (GO:0044763)	11573	1021	878.51	1.16	1.57E-09
single-organism process (GO:0044699)	12867	1107	976.74	1.13	3.41E-08
cellular process (GO:0009987)	14439	1211	1096.07	1.1	6.16E-07
biological regulation (GO:0065007)	11293	946	857.26	1.1	3.08E-02
biological_process (GO:0008150)	16739	1368	1270.66	1.08	5.53E-07
Unclassified (UNCLASSIFIED)	4075	213	309.34	0.69	0.00E+00

Table S5: Association tests of 49 SNVs with PBS>0.8 between IA cases and controls in Nunavik Inuit

CHR	SNP	BP	F_A	F_U	P	OR	GENE	FUNC	FCEU	FCHB	FGI	PBS
1	rs2276405	1.59E+08	0.2976	0.2177	0.1918	1.522	*AIM2*	missense	0.000505	0.029126	0.8611	1.279471
1	rs2070150	1.62E+08	0.1548	0.2097	0.3194	0.6901	*ATF6*	missense	0.045455	0.300971	#N/A	0.846154
1	rs1135983	1.62E+08	0.1548	0.2097	0.3194	0.6901	*ATF6*	missense	0.045455	0.291262	0.8889	0.862562
1	rs800292	1.97E+08	0.1071	0.05645	0.1782	2.006	*CFH*	missense	0.782828	0.635922	#N/A	0.891947
1	rs10864004	2.13E+08	0.02381	0.02419	0.9858	0.9837	*C1orf227*	missense	0.806122	0.553398	0.1944	1.031811
2	rs3827760	1.1E+08	0	0.008065	0.4093	0	*EDAR*	missense	1	0.067961	0.97222	1.348744
2	rs16866391	1.79E+08	0.1786	0.1129	0.1797	1.708	*TTN,TTN-AS1*	missense	0.000556	0.037313	#N/A	1.72467
3	rs180768267	12858557	0.1786	0.2339	0.338	0.7121	*CAND2*	missense	0.000556	0.021739	#N/A	1.389524
3	rs181307051	12861600	0.2619	0.3871	0.06081	0.5618	*CAND2*	missense	0.000538	0.000568	0.3889	1.03994
3	rs148960024	20189472	0.4643	0.3548	0.1137	1.576	*KAT2B*	missense	0.000505	0.009709	#N/A	0.882443
3	rs7614116	1.3E+08	0	0.01613	0.2422	0	*COL6A6*	missense	0.784884	0.439655	#N/A	0.98193
4	rs16851681	47578971	0.02381	0.01613	0.6923	1.488	*ATP10D*	missense	0.232323	0.194175	#N/A	1.433606
5	rs61739424	1.78E+08	0.3571	0.2984	0.3736	1.306	*COL23A1*	missense	0.042169	0.085714	#N/A	0.805394
5	rs634501	1.8E+08	0.04762	0.05645	0.78	0.8357	*MGAT1*	missense	0.787879	0.514563	#N/A	0.81513
6	rs1864312	49701439	0.02381	0.04032	0.517	0.5805	*CRISP3*	missense	0.60101	0.650485	#N/A	0.843453
7	rs2289367	45113170	0.07143	0.25	0.000952	0.2308	*CCM2*	synonymous-splicing	0.20202	0.194175	0.8889	0.820568
7	rs2304693	45148667	0.131	0.25	0.03585	0.4521	*TBRG4*	missense	0.171717	0.174757	0.8056	0.822423
7	rs2304694	45148773	0.131	0.25	0.03585	0.4521	*TBRG4*	missense	0.171717	0.174757	0.8056	0.822423
8	rs2741098	6690276	0.04762	0.02419	0.358	2.017	*XKR5*	missense	0.62069	0.572289	#N/A	0.811216
8	rs3764795	1.44E+08	0.03571	0.04098	0.8475	0.8667	*GML*	missense	0.954545	0.504854	#N/A	1.237918
8	rs4534	1.44E+08	0.03571	0.04032	0.8653	0.8815	*CYP11B1*	missense	0.015152	0.446602	#N/A	1.511838
9	rs78341003	1.07E+08	0.3286	0.4865	0.05412	0.5165	*OR13C5*	missense	0.000521	0.000617	#N/A	0.951218
10	rs117193245	38404188	0.3333	0.3629	0.6611	0.8778	*ZNF37A*	missense	0.005051	0.082524	0.5833	0.827687
10	rs3998860	70405855	0.08333	0.1774	0.05459	0.4215	*TET1*	missense	0.207071	0.165049	#N/A	0.984147
11	rs145672238	57080882	0.3452	0.3306	0.827	1.067	*TNKS1BP1*	missense	0.000505	0.004854	0.2222	1.063441
11	rs17152659	58170780	0.3571	0.2823	0.2529	1.413	*OR5B3*	missense	0.000505	0.082524	0.1389	0.936222

11	rs80356779	68548130	0.03571	0.04839	0.6594	0.7284	*CPT1A*	missense	0.000505	0.000485	#N/A	3.111911
11	rs560096	68678962	0.04762	0.04839	0.9797	0.9833	*IGHMBP2*	missense	0.883838	0.538835	0	1.028293
12	rs61729512	7637769	0.3333	0.3226	0.8711	1.05	*CD163*	missense	0.116162	0.014563	0.3889	0.812778
12	rs2066815	56743044	0.25	0.1774	0.2047	1.545	*STAT2*	missense-near-splice	0.000505	0.009709	#N/A	1.525998
13	rs7334118	52513266	0.3095	0.2823	0.6717	1.14	*ATP7B*	missense	0.005495	0.000588	0.4722	1.152985
13	rs17480245	52598189	0.3214	0.275	0.4738	1.249	*ALG11*	missense	0.010101	0.000485	0.4722	1.187379
13	rs7986131	76423248	0.04762	0.05645	0.78	0.8357	*LMO7*	missense	0.787879	0.548544	0.05556	0.865208
16	rs17856580	27246617	0.0119	0.008065	0.7806	1.482	*NSMCE1*	missense	0.773684	0.604167	0.8889	1.147446
16	rs1126464	89704365	0.08333	0.03226	0.1063	2.727	*DPEP1*	missense	0.70202	0.669903	0.75	0.924261
16	rs885479	89986154	0	0.02419	0.151	0	*MC1R*	missense	0.921348	0.436782	0.8889	1.199063
16	rs4264393	90130139	0.2262	0.1721	0.3351	1.406	*PRDM7*	missense	0.000602	0.007143	#N/A	1.504595
18	rs3744900	33779705	0.369	0.2823	0.187	1.487	*MOCOS*	missense	0.045455	0.053398	0.3889	0.916642
18	rs623053	33779855	0.369	0.2823	0.187	1.487	*MOCOS*	missense	0.045455	0.063107	0.6111	0.89798
18	rs678560	33785093	0.369	0.2661	0.1146	1.613	*MOCOS*	missense	0.045455	0.063107	0.6111	0.924335
18	rs2278154	72103782	0.119	0.06452	0.1699	1.959	*FAM69C*	missense	0.678571	0.8	0.8056	0.977466
19	rs5498	10395683	0.0119	0.02419	0.5266	0.4859	*ICAM1*	missense	0.585859	0.737864	0.8056	0.997435
19	rs1056538	10402938	0.0119	0.02419	0.5266	0.4859	*ICAM5*	missense	0.616162	0.81068	#N/A	1.163544
19	rs2228615	10403368	0.0119	0.02419	0.5266	0.4859	*ICAM5*	missense	0.616162	0.81068	0.8056	1.163544
19	rs1982074	10668673	0.1667	0.2177	0.3636	0.7185	*KRI1*	missense	0.151515	0.18932	#N/A	0.845432
19	rs2195980	44351694	0.02381	0.04839	0.3658	0.4797	*ZNF283*	missense	0.813131	0.601942	#N/A	1.048763
19	rs3764609	46823702	0.02381	0.04032	0.517	0.5805	*HIF3A*	missense	0.818182	0.538835	0.5278	0.979903
19	rs2278428	55418054	0.04762	0.04032	0.7997	1.19	*NCR1*	missense	0.065657	0.320388	#N/A	1.45223
22	rs2228314	42276742	0	0.008065	0.4093	0	*SREBF2*	missense	0.252525	0.194175	0.97222	1.48112

Appendix 3: Supplemental material for Chapter 2.3

Supplementary figures

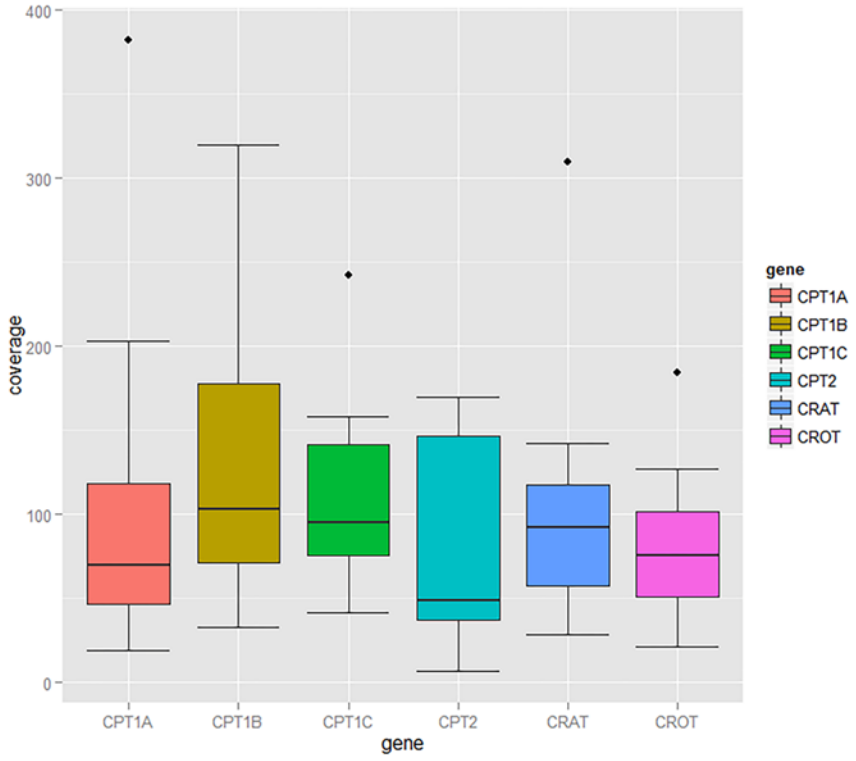

Figure S1: Average coverage for each exon of gene *CPT1A*, *CPT1B*, *CPT1C*, *CPT2*, *CRAT* and *CROT* from 113 Inuit exomes.

Figure S2(A-C): Chromatograms of Sanger sequencing validation of four rare missense variants and three rare silent variants found in Nunavik Inuit (*CPT1A* P479L from selected samples)

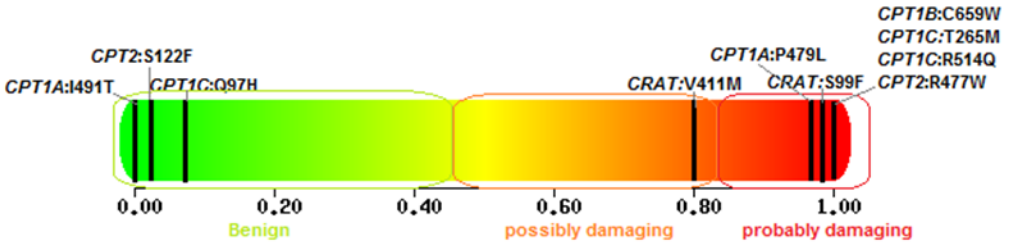

Figure S3: Rare variants of Nunavik Inuit and 1KGBP Asian and there gradation profile of Polyphen-2 scores.

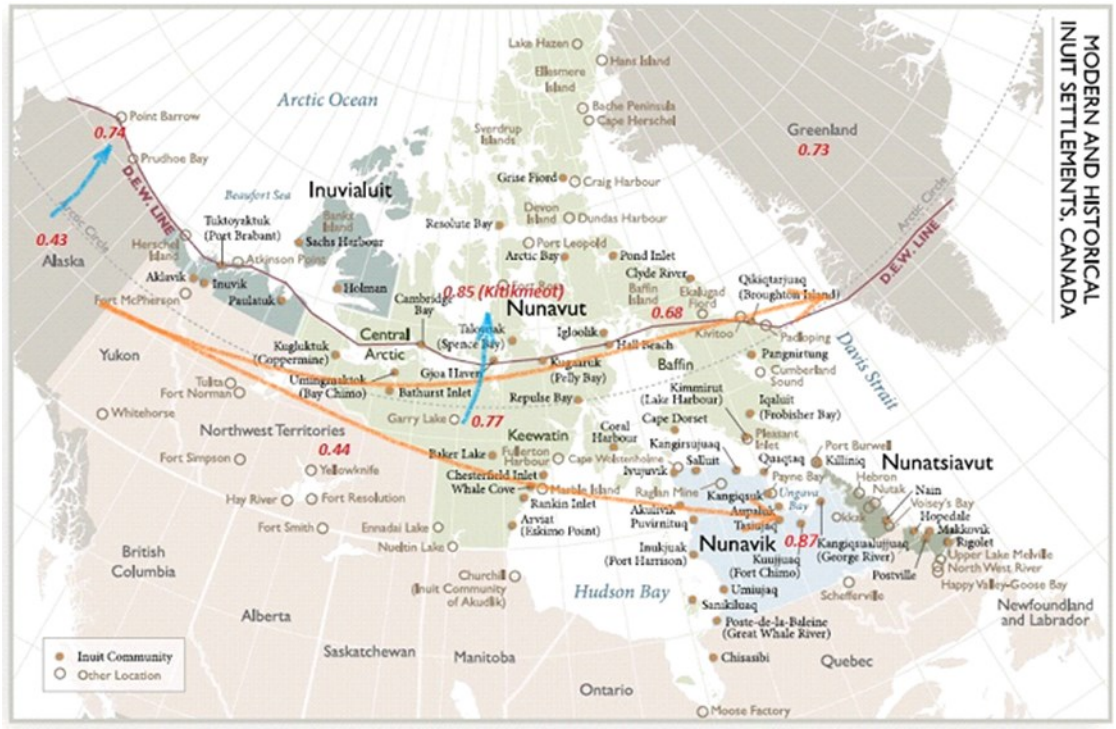

Figure S4: The CPT1A P479L variant frequency in different Inuit settlements in Canada and the migration patterns of Inuit.

Supplementary tables

Table S1: Variant concordance of *CPT1A, CPT1B, CPT1C, CPT2* and *CRAT* SNPs by genotyping and exome sequencing

Variant	**location**	**Gene**	**Success rate of genotyping (%)**	**Success rate of exome sequencing (%)**	**Variant concordance (%)**
rs3019598	Intronic	*CPT1A*	100	100	100
rs7112615	Intronic	*CPT1A*	100	100	100
rs2305508	Intronic	*CPT1A*	100	99	91
rs2924674	Intronic	*CPT1A*	100	93	100
rs7238	UTR-3	*CPT1B*	100	82	79
rs5770911	Intronic	*CPT1B*	100	100	100
rs12627787	Intronic	*CPT1B*	99	98	100
rs1557502	Intronic	*CPT1B*	100	43	65
rs3213445	Exonic	*CPT1B*	100	100	100
rs10407097	Intronic	*CPT1C*	100	35	80
rs3766759	Intronic	*CPT2*	100	15	40
rs12737375	Intronic	*CPT2*	100	80	86
rs1799821	Exonic	*CPT2*	100	100	100
rs1799822	Exonic	*CPT2*	100	100	100
rs3118635	Exonic	*CRAT*	100	100	100
rs10988209	Intronic	*CRAT*	100	45	73
			Mean	Mean	Mean
Exonic variant			100	100	100
Intronic variant			99.9	73.3	77

Table S2: All exonic variants of *CPT1A, CPT1B, CPT1C, CPT2* and *CART* in 100 Nunavik Inuit

gene	variant position (hg19)	variant function	nucleotide change	amino acid change	SNP	1KGP frequency	Nunavik Inuit frequency	Genotyped
CPT1A	chr11:68530122	Silent	c.G1848A	p.V616V	GCGACTTCGT[G/A]CGGGCCATGG	0	0.04	
CPT1A	chr11:68548130	Missense	c.C1436T	p.P479L	GCAGATGCGC[C/T]GATCGTGGCC	0	0.955	
CPT1A	chr11:68549340	Silent	c.T1251C	p.F417F	rs2228502	0.91	1	
CPT1B	chr22:51009953	Missense	c.G1591A (NM_152246)	p.E531K	rs470117	0.38	0.32	
CPT1B	chr22:51011348	Splicing	c.C1308T (NM_152246)	p.L436L	rs34744246	0.02	0	
CPT1B	chr22:51011376	Missense	c.C1280G (NM_152246)	p.S427C	rs8142477	0.36	0.66	
CPT1B	chr22:51015838	Missense	c.A196G (NM_152246)	p.I66V	rs3213445	0.17	0.185	yes
CPT1C	chr19:50208286	Missense	c.C794T	p.T265M	GTCACACCCA[C/T]GCCTCTGCAG	0	0.005	
CPT2	chr1:53676401	Missense	c.T1055G	p.F352C	rs2229291	0.06	0.275	
CPT2	chr1:53676448	Missense	c.G1102A	p.V368I	rs1799821	0.5	0.535	yes
CPT2	chr1:53676775	Missense	c.C1429T	p.R477W	GGCCTTCCTG[C/T]GGCAGTACGG	0	0.03	
CPT2	chr1:53679229	Missense	c.A1939G	p.M647V	rs1799822	0.13	0	yes
CRAT	chr9:131857687	Missense	c.G1807C (NM_000755)	p.A603P	rs17459086	0.04	0.1	yes
CRAT	chr9:131857769	Silent	c.G1725A (NM_000755)	p.A575A	ACAGCTGCGC[G/A]GAGACCAACG	0	0.02	
CRAT	chr9:131860901	Missense	c.C1051A (NM_000755)	p.L351M	rs3118635	1	1	yes
CRAT	chr9:131866581	Splicing	c.C233T (NM_000755)	p.S78F	TCCCAGCTGT[C/T]TGAGTGGTGG	0	0.02	

Appendix 4: Supplemental material for Chapter 3.2

Supplementary figures

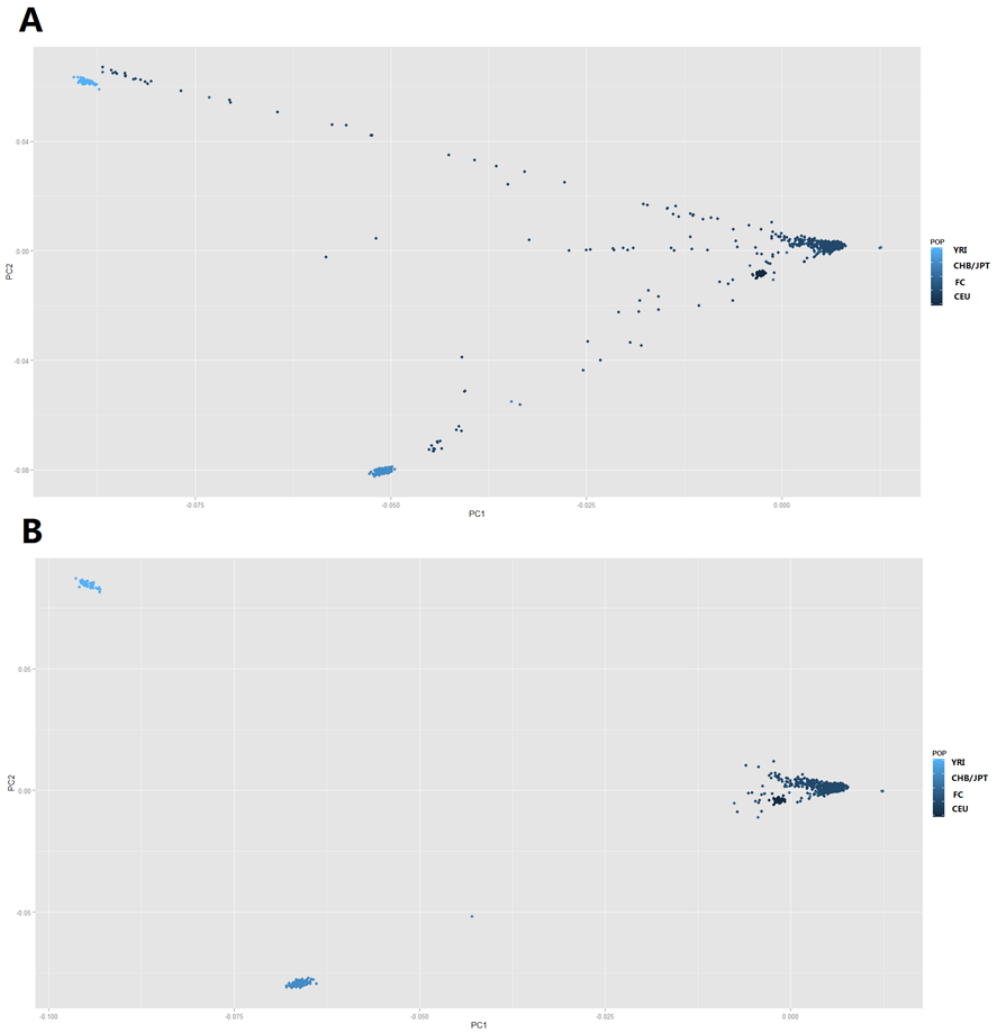

Figure S1: PCA analysis of FC patients and controls

PCA analysis done using smartPCA and graph constructed by R. (A) PCA plot of the FC NeuroX cohort (B) Samples involved in this study after the removal of 95 admixed individuals. YRI, CHB/JPT and CEU populations from 1KG phase III were used for clustering and outliers with less than 90% of European ancestry are removed.

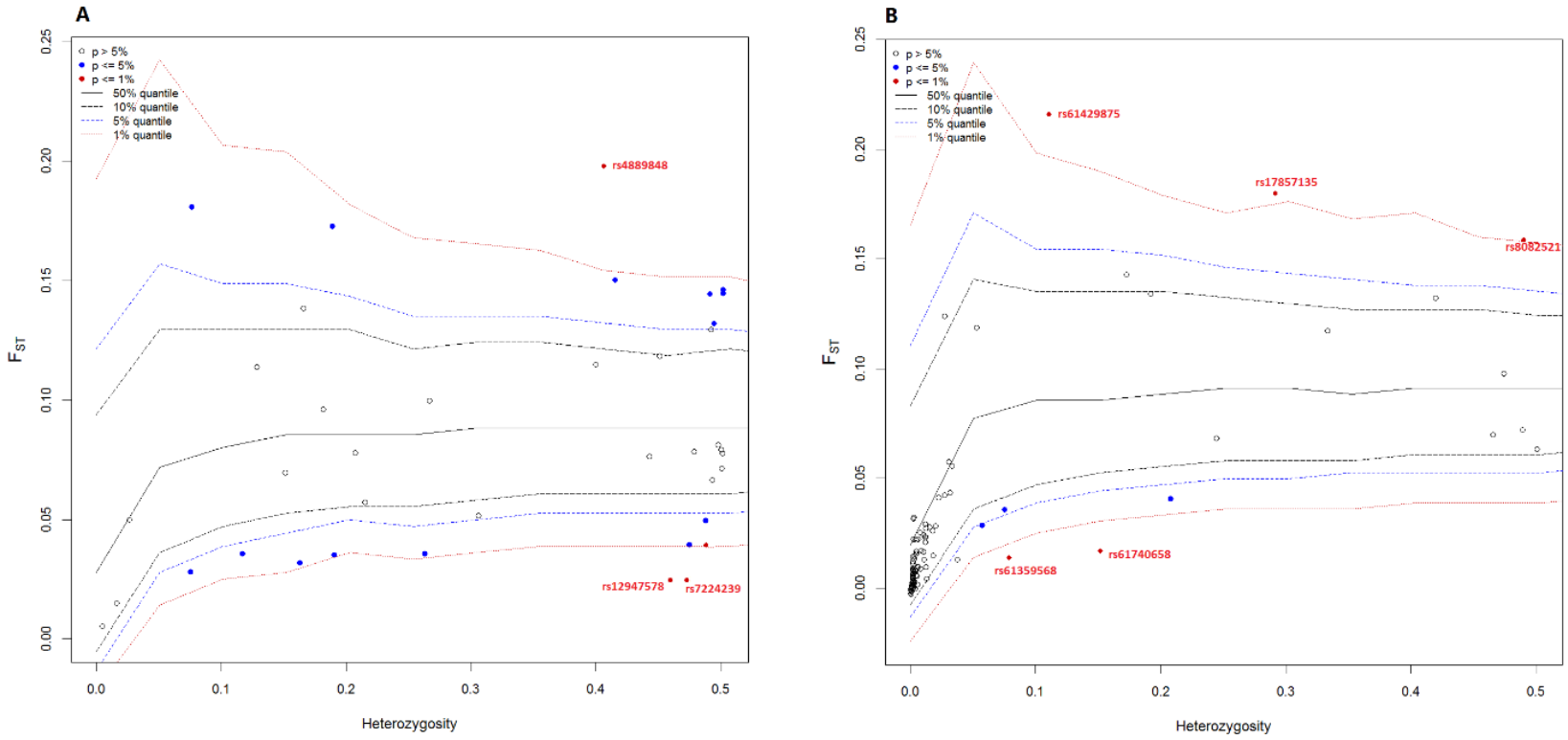

Figure S2: F_{ST} of *RNF213* common variations between worldwide populations

F_{ST} of *RNF213* common variants (A) and (B) functional variants, SNPs in red indicate they are outliers of 50% quantile.

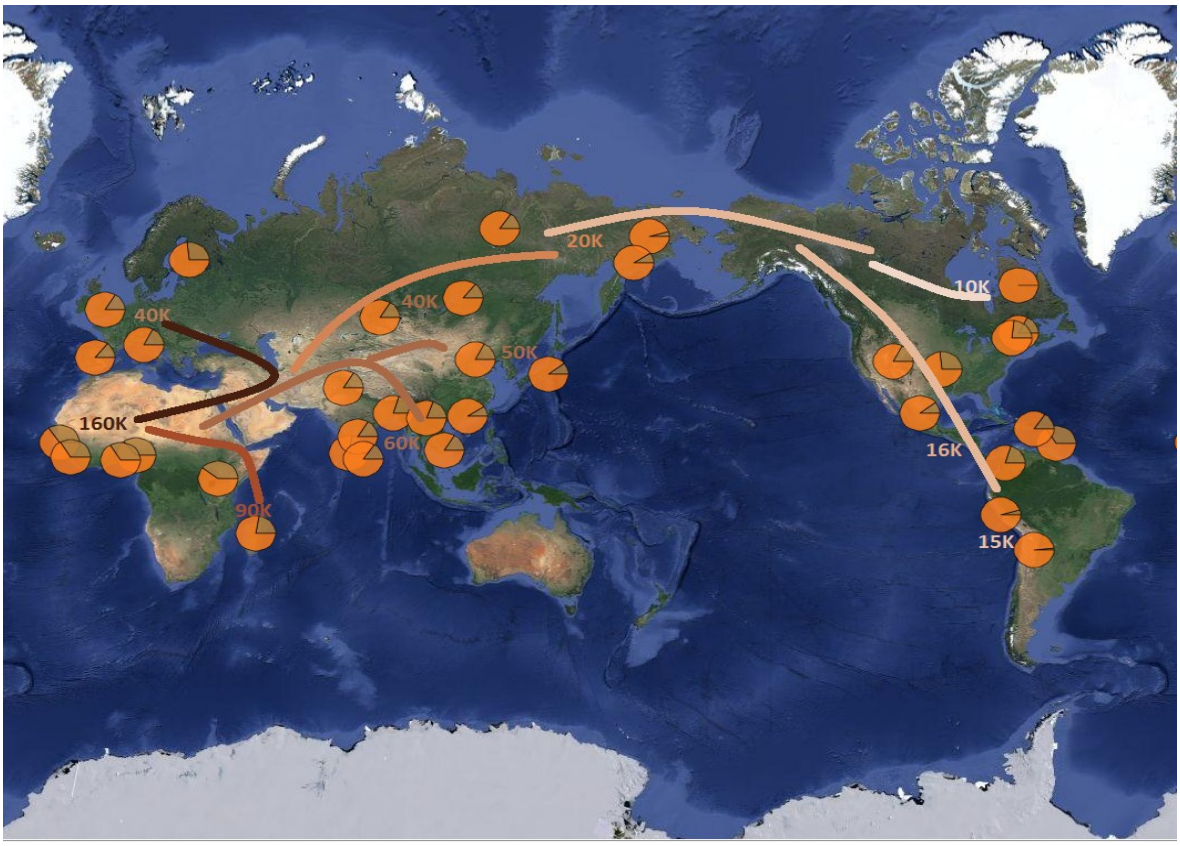

Figure S3: MAF of *RNF213* SNP rs6565666 distribution in world-wide populations

The frequency of rs6565666 non-ancestral allele decreased from Africans (0.26-0.35) to Europeans (0.16-0.18) and Asians (0.17-0.1) to the lowest in arctic populations (0.04-0.003). This non-ancestral allele seems also to be the risk allele for IA (dominant model for minor allele, p=0.0001739), which seems to correlate with the Out-of-Africa migration patterns and risk allele being purged in the evolution.

Supplementary tables

Table S1. Clinical descriptions of the IA discovery cohort and replication cohort.

	Number of individuals in discovery cohort	Number of individuals in replication cohort (familial)	Number of individuals in replication cohort (sporadic)
With clinical file	26	147	65
Multiple aneurysms	12	55	19
SAH	5	51	25
Hypertension	4	60	31
Drinking	12	33	26
Smoking	24	89	50
Hypercholesterolemia	3	29	24
Hypothyroidism	0	17	14
Had heart diseases and other vascular diseases	0	13	3

Table S3: Population data retrieved for this study

Region	Populations	No. Samples	Data access	Data source	SNPs in RNF213
Argentina	Argentina	50	GEO	Omni	40
Siberia	Chukotka	48	GEO	Omni	48
Siberia	East-Siberia	85	GEO	Omni	48
Kamchatka Krai	Kamchatka Krai	67	GEO	Omni	48
Sakha Republic	Sakha Republic	46	GEO	Omni	48
West-Siberia	West-Siberia	103	GEO	Omni	48
Africa	Madagascar	69	GEO	Omni	39
East_Asia	Han Chinese	203	1000 Genome	Omni & Exome	1835
East_Asia	JPT	104	1000 Genome	Omni & Exome	1835

East_Asia	CDX	93	1000 Genome	Omni & Exome	1835
East_Asia	Vietnamese	117	1000 Genome & GEO	Omni & Exome	1835
Europe	CEU	99	1000 Genome	Omni & Exome	1835
Europe	TSI	107	1000 Genome	Omni & Exome	1835
Europe	GBR	91	1000 Genome	Omni & Exome	1835
Europe	FIN	99	1000 Genome	Omni & Exome	1835
Europe	IBS	107	1000 Genome	Omni & Exome	1835
Africa	YRI	108	1000 Genome	Omni & Exome	1835
Africa	LWK	99	1000 Genome	Omni & Exome	1835
Africa	GWD	113	1000 Genome	Omni & Exome	1835
Africa	MSL	85	1000 Genome	Omni & Exome	1835
Africa	ESN	99	1000 Genome	Omni & Exome	1835
Native-America	ASW	61	1000 Genome	Omni & Exome	1835
Native-America	ACB	96	1000 Genome	Omni & Exome	1835
Native-America	MXL	64	1000 Genome	Omni & Exome	1835
Native-America	PUR	104	1000 Genome	Omni & Exome	1835
Native-America	CLM	94	1000 Genome	Omni & Exome	1835
Native-America	PEL	85	1000 Genome	Omni & Exome	1835
South_Asia	GIH	103	1000 Genome	Omni & Exome	1835
South_Asia	PJL	96	1000 Genome	Omni & Exome	1835
South_Asia	BEB	86	1000 Genome	Omni & Exome	1835
South_Asia	STU	102	1000 Genome	Omni & Exome	1835
South_Asia	ITU	102	1000 Genome	Omni & Exome	1835
Inuit	Inuit	185	In house & GEO	Omni	39

Table S4: Exome sequencing coverage of 26 IA patients from the initial cohort

Indiv	Raw reads	% Coverage (10X)	20X	30X	50X	75X	100X	Average base depth
09-1	7.90E+07	97.1	90.7	81.2	59.8	35.5	18.5	66.07
09-2	8.80E+07	97.8	93.6	86.5	68.4	46.1	28.1	78.44
28-1	1.40E+08	98.4	97	94.7	87.5	75.8	63.4	144.4
28-2	1.40E+08	98.7	97.1	94.3	85.8	72.8	59.5	133.63
28-3	1.04E+08	98.2	95.6	90.9	77.7	59.5	42.6	99.55
28-4	9.10E+07	98.2	95	89.5	75	55.8	38.7	93.58
94-1	1.10E+08	98.5	96.7	93.3	82.4	65.7	49.7	111.25
94-2	1.10E+08	98.3	95.7	91.2	79.6	63.3	47.4	108.05
94-3	1.30E+08	98.8	97.3	94.9	87	74.3	61	137.29
94-4	1.10E+08	98.5	96.7	93.4	82.5	65.9	50	112.03
60-1	2.60E+09	90.9	85.5	79.3	62.6	40.9	25.8	78.36
60-2	3.10E+09	91.9	87.4	82.6	69.2	49.6	33.7	92.27
60-3	2.70E+09	91.5	86.6	80.9	64.9	43.2	27.7	81.79
60-4	3.40E+09	91.8	87.6	83.3	72.6	55.5	39.6	102.47
60-5	2.70E+09	90.6	85.3	79.2	63.4	42.5	27.4	80.68
10-1	3.10E+09	91.5	86.8	81.9	68.9	49.4	33.4	91.2
10-2	3.20E+09	91.7	87.2	82.7	70.9	52.8	36.7	97
10-3	3.80E+09	92.8	89.1	85.3	75.5	59.7	44.5	113.59
10-4	3.60E+09	92.5	87	81.2	69.4	54.7	41.3	108.23
10-5	3.10E+10	91.4	86.9	82	68.8	49.1	33.2	91.07
89-1	6.00E+09	93.6	90.6	88	83.2	75.9	67.1	177.72
89-2	1.20E+09	82.6	65.5	47	22.3	9.5	4.5	35.36
89-3	2.60E+09	90.2	84.2	77.7	61	39.7	25.1	76.43
89-4	2.60E+09	90.2	84.4	78	61.9	41.1	26.2	78.39
89-5	3.40E+09	91.5	87.1	82.7	71.7	55	39.6	102.88
89-6	4.20E+09	93	89.3	86.1	78.6	66	51.7	126.19

Table S5: 72 RNF213 functional variants in FC IA patients and controls

Position	function	Exon	Annotation	dbSNP v132	1000 Genome	NHLBI-ESP	PolyPhen2	Case maf(233)	Control maf(277)
78247087	missense	3	c.A145G:p.M49V	rs114493717	0.01	0.005152	0.400054	0.002146	0
78261704	missense	4	c.T352C:p.C118R	NA	0.0005	0.000154	0.266834	0.008584	0.001805
78261961	missense	4	c.G609C:p.E203D	rs74003706	0.02	0.019453	0.118686	0.002146	0
78262538	missense	5	c.T917C:p.F306S	NA	0.0014	0.000461	0	0	0.001805
78263486	missense	6	c.T962C:p.M321T	rs17853989	0.2	0.259111	0.006	0.173809	0.1498
78264463	nonframeshift_ deletion	7	c.1208_1210del:p.403_404del	NA	0.0023	0.001917	1	0.002146	0
78265614	missense	8	c.C1459A:p.L487M	NA	0	0	0.387	0.002146	0.001805
78268746	missense	9	c.A1699G:p.M567V	NA	0	0	0.999	0.002146	0
78272125	splicing	11	c.C2017T:p.R673W	NA	0	0	0.997	0.006438	0
78272230	missense	11	c.C2122G:p.H708D	rs72849837	0.01	0.012917	0.181	0.012874	0.01264
78272294	missense	11	c.C2186T:p.P729L	rs72849841	0.11	0.145241	0.026	0.145955	0.1296
78291060	missense	16	c.G2884A:p.E962K	NA	0	0	0.068	0	0.001805
78293189	missense	17	c.A3101T:p.K1034M	rs55996424	0.34	0.189451	0.541746	0.160946	0.1667
78293222	missense	17	c.C3134T:p.S1045L	NA	0	0	0.496412	0.002146	0.005415
78298872	missense	18	c.C3067T:p.R1023W	NA	0	0	0.001	0.002146	0
78298926	missense	18	c.G3121A:p.A1041T	rs61359568	0.04	0.051029	0.007	0.072957	0.04815

78302157	missense	20	c.C3397A:p.Q1133K	rs8082521	0.56	0.700832	0	0.270336	0.3315
78302164	missense	20	c.C3404T:p.A1135V	NA	0	0	0.223	0	0.00722
78305871	missense	21	c.G3583A:p.V1195M	rs10782008	0.52	0.59067	0.73	0.329691	0.3662
78305896	missense	21	c.C3608T:p.S1203L	NA	0	0	0.154	0	0.00361
78305962	missense	21	c.A3674G:p.D1225G	NA	0	0	0.919	0.002146	0
78305989	missense	21	c.A3701G:p.Q1234R	rs116722283	0.02	0	0.493	0.002146	0
78306102	missense	21	c.G3814C:p.E1272Q	rs9913636	0.37	0.386553	0.69	0.49327	0.469
78306107	missense	21	c.G3819C:p.E1273D	NA	0	0	0	0	0.001805
78306280	missense	21	c.A3992G:p.D1331G	rs8074015	0.6	0.703241	0.002	0.273071	0.3259
78307917	missense	22	c.G4156A:p.D1386N	rs113139767	0.0023	0.002847	0.154	0	0.001805
78310056	missense	23	c.G4405C:p.V1469L	rs116694967	0.01	0.010293	0.908	0.002146	0
78313074	missense	26	c.C4907T:p.T1636M	NA	0.0023	0	0.96	0.008584	0.00361
78313281	missense	26	c.C5114A:p.T1705K	NA	0.0032	0.007008	1	0.025747	0.007246
78313329	missense	26	c.C5162T:p.P1721L	NA	0.0014	0	0.877	0.002146	0.00361
78313622	missense	26	c.C5455T:p.R1819C	NA	0	0	0.999	0	0.001805
78313697	missense	26	c.G5530A:p.A1844T	NA	0	0	0.041	0.004292	0.005415
78313763	missense	26	c.A5596G:p.T1866A	NA	0	0	0.77	0.006438	0
78316974	missense	27	c.T6032C:p.M2011T	rs78457838	0.01	0.011841	0.004	0.002146	0
78318782	missense	29	c.G6647A:p.R2216Q	NA	0	0	0.66	0.002146	0
78319114	missense	29	c.A6979G:p.N2327D	NA	0.0009	0.000923	0.049	0.004292	0

78319115	missense	29	c.A6980G:p.N2327S	NA	0	0	0.094	0.002146	0.001805
78319136	missense	29	c.G7001A:p.S2334N	rs9674961	0.57	0.660464	0	0.300419	0.3466
78319447	missense	29	c.C7312T:p.R2438C	NA	0	0.000077	0.655	0.002146	0
78320611	missense	29	c.G8476A:p.A2826T	NA	0	0	1	0.002146	0
78320710	missense	29	c.A8575G:p.T2859A	NA	0	0	0.012	0	0.001805
78321046	missense	29	c.A8911G:p.N2971D	NA	0	0	0.012	0	0.00361
78321108	missense	29	c.T8973G:p.D2991E	NA	0	0.000231	0.528	0.002146	0.001805
78321157	missense	29	c.T9022C:p.C3008R	rs61600413	0.08	0.113409	0.803	0.062231	0.03889
78321560	missense	29	c.T9425C:p.V3142A	NA	0	0	0.998	0	0.001805
78321697	missense	29	c.G9562A:p.V3188M	NA	0	0	0.96	0.002146	0
78321844	missense	29	c.C9709A:p.Q3237K	NA	0	0	0.264	0.006438	0
78321941	missense	29	c.G9806A:p.R3269Q	NA	0	0.000077	1	0	0.00361
78323570	splicing	30	c.A9952G:p.I3318V	NA	0.0018	0.002153	1	0.002146	0.001805
78328364	missense	36	c.C10850T:p.A3617V	NA	0	0	0.555	0	0.001805
78336958	frameshift_deletion	40	c.11413delC:p.P3805fs	NA	0	0	1	0.004292	0.001805
78337058	missense	40	c.G11512C:p.V3838L	rs35332090	0.13	0.154698	0.329	0.145955	0.1029
78337440	missense	41	c.C11600T:p.T3867M	NA	0	0.000154	0.504	0	0.005415
78337584	missense	41	c.A11744G:p.E3915G	rs61740658	0.07	0.085807	0.243	0.111619	0.07581
78343401	missense	45	c.G12255C:p.E4085D	NA	0	0	0	0	0.001805

78345744	missense	47	c.G12496A:p.D4166N	NA	0	0.000923	1	0.002146	0
78346840	missense	49	c.G12817A:p.D4273N	NA	0.0023	0.003537	1	0.008584	0.009025
78346870	missense	49	c.C12847A:p.L4283I	rs62077764	0.04	0.05436	0.264	0.051505	0.06137
78346877	missense	49	c.G12854A:p.S4285N	NA	0	0.000077	0.024	0.002146	0
78348389	splicing	50	c.G13074A:p.K4358K	NA	0	0	1	0.002146	0
78350110	missense	52	c.G13195A:p.A4399T	NA	0.02	0.006689	0.906	0.004292	0.009025
78350767	splicing	53	c.13510+4A>G	NA	0.0009	0.004998	1	0.010726	0.005415
78353451	missense	55	c.T13577C:p.I4526T	NA	0	0	0.290848	0.006438	0
78354656	missense	56	c.G13666A:p.G4556S	NA	0	0	0.998	0.002146	0.00361
78355462	missense	57	c.C13913T:p.T4638I	NA	0	0.001	0.995	0.006438	0.001805
78355494	missense	57	c.C13945G:p.L4649V	rs61745599	0.05	0.074273	0.996	0.075105	0.08845
78356830	missense	58	c.G14030T:p.W4677L	rs61741961	0.01	0.014301	0.997	0.006438	0.02347
78357478	splicing	59	c.A14072G:p.H4691R	rs12944088	0.02	0.019914	1	0.036484	0.01625
78357600	missense	59	c.A14194G:p.K4732E	rs12944385	0.0032	0.007766	0.087	0.006438	0.005415
78363054	missense	65	c.G15082A:p.V5028I	rs8072774	0.1	0.156005	0.011	0.079399	0.08664
78363707	missense	66	c.G15275A:p.R5092Q	NA	0	0.000077	0.998	0.002146	0
78367225	missense	68	c.A15551C:p.E5184A	rs61429875	0.05	0.069276	0.734	0.002146	0

Table S6. Cochran-Armitage trend test on 38 polymorphic SNPs in *RNF213* loci

CHR	SNP	A1	A2	AFF	UNAFF	CHISQ	DF	P
17	rs7217421	A	G	248/256	1722/1996	1.478	1	0.2241
17	rs4889968	G	A	12/484	117/3513	0.8884	1	0.3459
17	rs9916351	C	T	221/287	1757/1969	2.436	1	0.1186
17	rs12947578	A	G	202/306	1583/2139	1.444	1	0.2295
17	rs9913317	A	G	0/502	1/3733	0.1345	1	0.7138
17	rs7219131	C	T	224/280	1780/1940	2.13	1	0.1444
17	rs12601730	G	T	12/496	113/3615	0.6745	1	0.4115
17	rs7220465	T	C	128/374	1092/2636	3.117	1	0.0775
17	rs11869363	C	T	136/370	1173/2545	4.732	1	0.0296
17	rs12937242	T	C	77/431	687/3037	3.324	1	0.06827
17	rs8068939	T	G	125/383	1004/2728	1.193	1	0.2747
17	rs12451223	C	T	120/388	935/2795	0.501	1	0.4791
17	rs8081176	C	T	159/349	1265/2467	1.327	1	0.2494
17	rs6565666	A	G	119/389	649/3085	11.29	1	0.00078
17	rs7225029	A	G	45/461	280/3450	1.255	1	0.2626
17	rs4889843	G	A	246/262	1681/2045	1.996	1	0.1577
17	rs9907978	A	G	146/362	1243/2487	4.172	1	0.0411
17	rs9908583	G	A	143/359	1232/2456	4.754	1	0.02923
17	rs4890008	T	G	27/481	151/3577	1.791	1	0.1807
17	rs10782008	G	A	173/335	1393/2331	2.139	1	0.1436
17	rs8072917	C	T	173/335	1393/2331	2.148	1	0.1428
17	rs4890009	G	A	144/358	1242/2484	4.266	1	0.03889
17	rs4890010	C	T	173/335	1393/2327	2.209	1	0.1372
17	NeuroX_chr17:78319717	T	C	0/508	2/3732	0.2724	1	0.6018
17	rs11150856	C	T	68/440	544/3188	0.4948	1	0.4818
17	rs12051723	T	G	98/410	545/3183	7.603	1	0.005829
17	rs11655038	T	G	40/468	337/3391	0.7355	1	0.3911
17	rs7216493	G	A	128/380	1086/2642	3.345	1	0.0674
17	rs8078251	G	A	170/338	1230/2490	0.03161	1	0.8589
17	rs4078429	G	A	28/478	198/3532	0.04479	1	0.8324
17	rs6565681	A	G	76/430	610/3114	0.5957	1	0.4402
17	rs4889848	C	T	75/433	608/3124	0.7565	1	0.3844
17	rs7224239	G	A	249/259	1622/2112	5.759	1	0.01641
17	rs12944385	G	A	3/505	22/3712	1.44E-05	1	0.997
17	rs4890018	C	T	31/477	210/3516	0.1718	1	0.6785
17	rs8072774	T	C	40/466	324/3406	0.3454	1	0.5567
17	rs3185057	A	G	51/457	312/3418	1.541	1	0.2144
17	rs8359	T	C	42/466	332/3400	0.2201	1	0.6389

Appendix 5: Supplemental material for Chapter 3.3

Supplementary figures

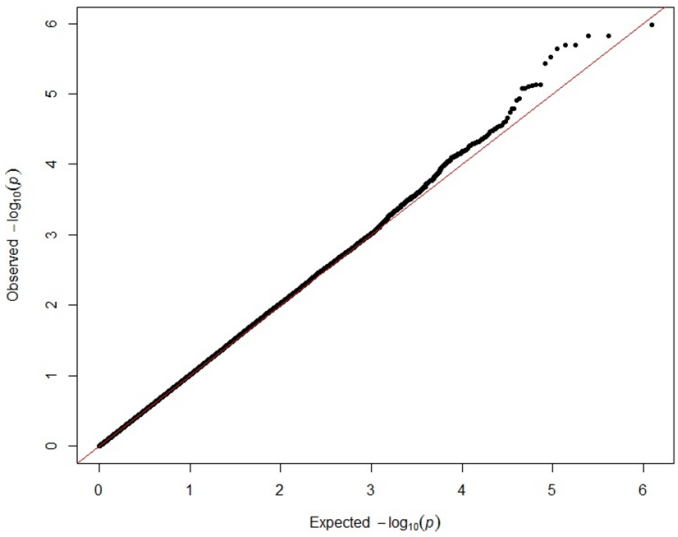

Figure S1. Quantile-quantile plot of case-control logistic regression analysis

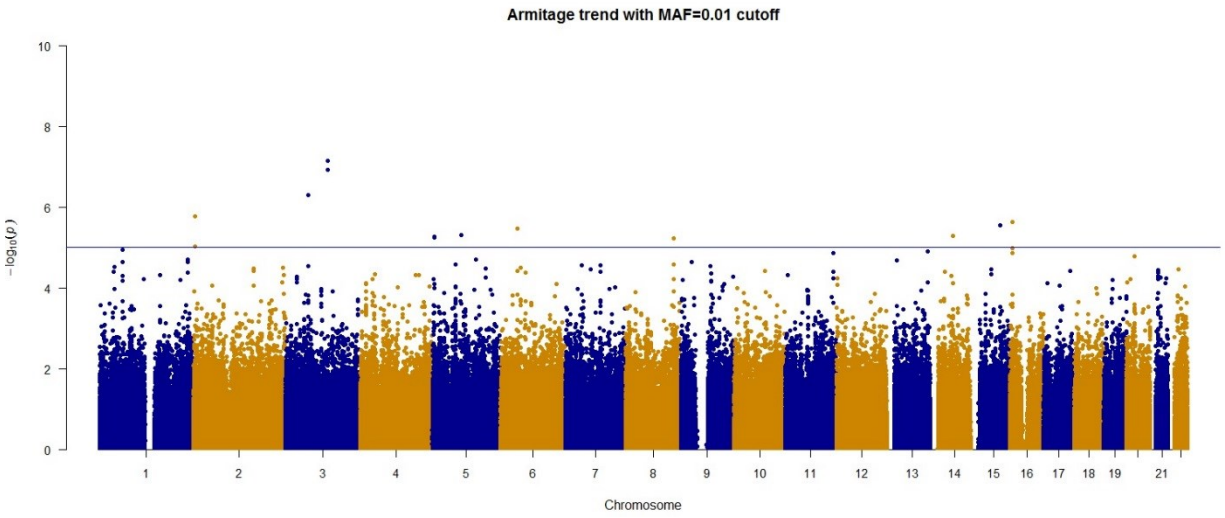

Figure S2. Manhattan plot showing the result of the initial trend test

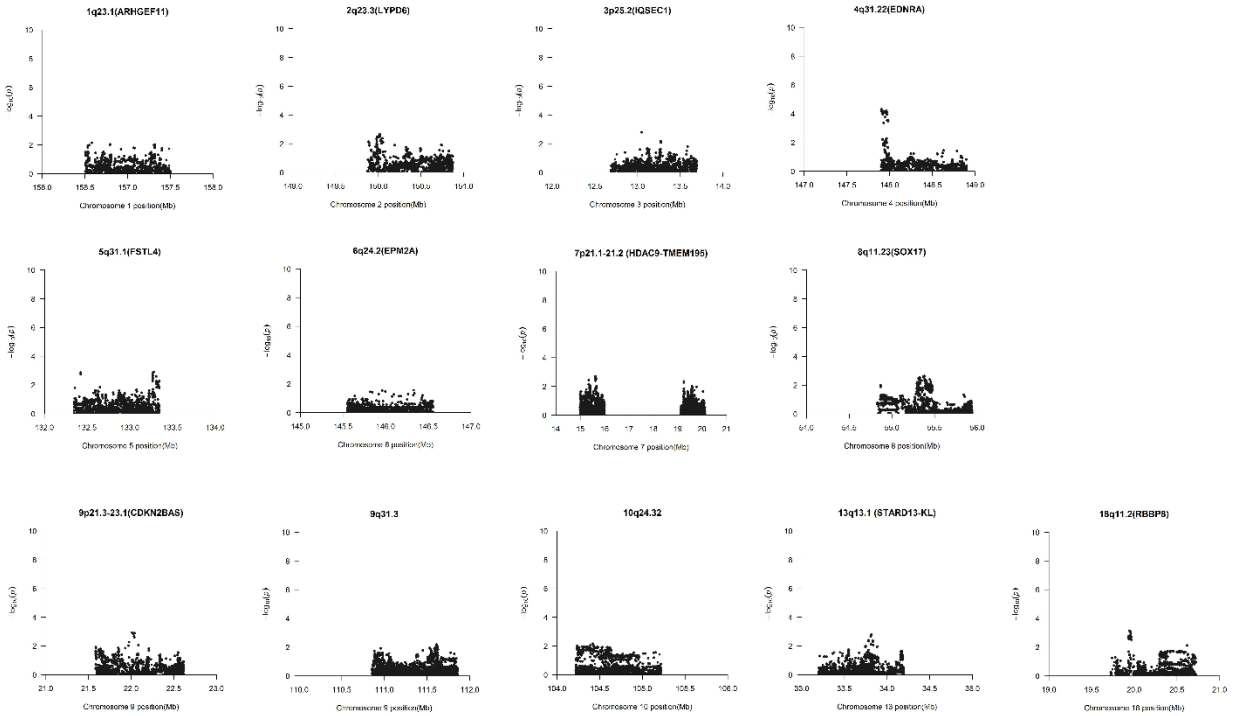

Figure S3. Regional validation of 13 previous GWAS associated loci

Supplementary tables

Table S1. FBAT-O result of 50 *FHIT* variants with significant association ($p<0.05$) in Nunavik Inuit

Marker	chr	pos	afreq	fam#	S-E(S)	Var(S)	Z	P	Offset
rs780365	3	59761789	0.377	14	6.842	5.254	2.985	0.002839	0.231
rs17361653	3	59771515	0.462	18	-6.903	5.489	-2.946	0.003216	0.203
rs9809480	3	60448671	0.23	14	5.032	3.155	2.833	0.004615	0.277
rs1597104	3	60453227	0.23	14	5.032	3.155	2.833	0.004615	0.277
rs7621114	3	60418721	0.205	13	4.663	3.003	2.691	0.007123	0.257
rs2255772	3	60557852	0.207	10	4.039	2.519	2.545	0.010928	0.245
rs2687189	3	60423009	0.438	18	4.701	3.768	2.422	0.015436	0.095
rs445057	3	60447369	0.433	18	4.701	3.768	2.422	0.015436	0.095
rs424675	3	60448261	0.433	18	4.701	3.768	2.422	0.015436	0.095
rs17061176	3	59755841	0.179	11	4.311	3.188	2.414	0.015763	0.224
rs13092167	3	59772342	0.293	10	4.244	3.148	2.392	0.016762	0.212
rs3920477	3	60733152	0.328	15	4.425	3.484	2.371	0.017762	0.154
rs9861525	3	60004908	0.398	15	4.756	4.157	2.332	0.019677	0.268
rs241700	3	60516256	0.311	13	4.01	3.005	2.313	0.020728	0.235
rs10451993	3	60566492	0.321	13	4.01	3.005	2.313	0.020728	0.235
rs17669257	3	60566161	0.412	17	4.626	4.158	2.269	0.023295	0.271
rs17609699	3	60487578	0.48	18	-5.16	5.188	-2.265	0.023496	0.27
rs49413	3	59757484	0.418	20	5.311	5.546	2.255	0.024112	0.264
rs993000	3	59745585	0.158	10	2.989	1.815	2.218	0.026535	0.4
rs1683346	3	59749703	0.158	10	2.989	1.815	2.218	0.026535	0.4
rs17061173	3	59754178	0.156	10	2.989	1.815	2.218	0.026535	0.4
rs994931	3	59741758	0.319	15	-4.661	4.506	-2.196	0.028127	0.451
rs780366	3	59762116	0.347	15	-4.661	4.506	-2.196	0.028127	0.451
rs11130733	3	59767786	0.347	15	-4.661	4.506	-2.196	0.028127	0.451
rs17361780	3	59772328	0.347	15	-4.661	4.506	-2.196	0.028127	0.451
rs1627302	3	60658608	0.473	17	-4.48	4.468	-2.119	0.034058	0.27
rs721207	3	60607896	0.468	16	-5.031	5.672	-2.112	0.034649	0.372
rs780354	3	59772460	0.366	15	-4.505	4.578	-2.106	0.035223	0.478
rs2682931	3	60631936	0.359	19	-4.939	5.68	-2.072	0.038225	0.261
rs6766789	3	60090013	0.468	20	-4.352	4.463	-2.06	0.039397	0.349
rs10780042	3	60090402	0.468	20	-4.352	4.463	-2.06	0.039397	0.349

rs9881736	3	60487247	0.139	11	3.056	2.21	2.055	0.039836	0.5
rs6786100	3	60351746	0.402	17	4.528	5.083	2.008	0.04461	0.302
rs2142301	3	60095070	0.396	19	-3.978	4.074	-1.971	0.048745	0.18
rs2885865	3	60610341	0.381	19	-4.741	5.811	-1.967	0.049201	0.292
rs2363670	3	60611613	0.372	19	-4.741	5.811	-1.967	0.049201	0.292
rs7633853	3	60613659	0.38	19	-4.741	5.811	-1.967	0.049201	0.292
rs11926787	3	60615310	0.368	19	-4.741	5.811	-1.967	0.049201	0.292
rs6762641	3	60616153	0.356	19	-4.741	5.811	-1.967	0.049201	0.292
rs10510852	3	60616790	0.356	19	-4.741	5.811	-1.967	0.049201	0.292
rs12637393	3	60620284	0.368	19	-4.741	5.811	-1.967	0.049201	0.292
rs2856058	3	60625598	0.36	19	-4.741	5.811	-1.967	0.049201	0.292
rs1735448	3	60647540	0.36	19	-4.741	5.811	-1.967	0.049201	0.292
rs1716709	3	60657213	0.349	19	-4.741	5.811	-1.967	0.049201	0.292
rs1716714	3	60660179	0.349	19	-4.741	5.811	-1.967	0.049201	0.292
rs1735443	3	60662282	0.349	19	-4.741	5.811	-1.967	0.049201	0.292
rs1735444	3	60664075	0.348	19	-4.741	5.811	-1.967	0.049201	0.292
rs1716722	3	60681361	0.349	19	-4.741	5.811	-1.967	0.049201	0.292
rs1735457	3	60683737	0.349	19	-4.741	5.811	-1.967	0.049201	0.292
rs1735468	3	60703616	0.36	19	-4.741	5.811	-1.967	0.049201	0.292

Table S2. 68 genes with exonic variants of FC exome cohort in 13 previously GWAS identified loci

Loci	**Gene**	**Number of exonic variants**	**MAC in each gene**	***P*-value**	**std error (SE)**	**permutations**
1q23.1	*IQGAP3*	10	219	0.516484	0.376007	1000
1q23.1	*TTC24*	9	123	0.380619	0.475192	1000
1q23.1	*APOA1BP*	6	62	0.0527894	0.583171	5000
1q23.1	*GPATCH4*	5	48	0.064787	0.521149	5000
1q23.1	*BCAN*	9	131	0.979021	0.379929	1000
1q23.1	*NES*	15	643	0.15984	0.395594	1000
1q23.1	*ISG20L2*	3	16	0.576424	0.385064	1000
1q23.1	*RRNAD1*	3	4	0.505495	0.432044	1000
1q23.1	*PRCC*	4	66	0.00359928	0.368741	5000
1q23.1	*SH2D2A*	3	104	0.300699	0.372431	1000
1q23.1	*INSRR*	18	72	0.0115977	0.408737	5000
1q23.1	*NTRK1*	13	207	0.495504	0.550519	1000

1q23.1	*PEAR1*	19	479	0.147852	0.539598	1000
1q23.1	*LRRC71*	8	269	0.253746	0.404279	1000
1q23.1	*ARHGEF11*	10	294	0.297702	0.398201	1000
1q23.1	*ETV3L*	10	261	0.873127	0.419813	1000
1q23.1	*ETV3*	2	29	0.127872	0.398124	1000
1q23.1	*FCRL5*	6	22	0.0677864	0.501448	5000
2q23.3	*MMADHC*	3	42	0.293706	0.411939	1000
3p25.2	*CAND2*	25	832	0.330669	0.37951	1000
3p25.2	*RPL32*	2	6	0.906094	0.421056	1000
3p25.2	*IQSEC1*	15	60	1	0.412235	1000
3p25.2	*NUP210*	25	1325	0.612388	0.359059	1000
3p25.2	*FBLN2*	28	584	0.214785	0.615247	1000
4q31.22	*EDNRA*	2	132	0.994006	0.486694	1000
4q31.22	*ARHGAP10*	6	9	0.956044	0.408415	1000
4q31.22	*PRMT10*	2	10	0.618382	0.437471	1000
5q31.1	*HSPA4*	7	95	0.158841	0.361348	1000
5q31.1	*FSTL4*	7	100	0.873127	0.363665	1000
6q24.2	*EPM2A*	7	183	0.297702	0.338059	1000
6q24.2	*FBXO30*	8	250	0.0103979	0.361532	5000
6q24.2	*SHPRH*	8	60	0.608392	0.391281	1000
7p21.2	*AGMO*	8	19	0.222777	0.434542	1000
7p21.2	*MEOX2*	2	4	0.270729	0.410235	1000
7p21.2	*FERD3L*	2	72	0.539461	0.386999	1000
7p21.2	*TWISTNB*	3	11	0.0641872	0.406677	5000
8q11.23	*RGS20*	2	2	0.724276	0.401227	1000
8q11.23	*LYPLA1*	4	5	0.872128	0.433138	1000
8q11.23	*SOX17*	3	5	0.0825835	0.409256	5000
8q11.23	*RP1*	17	614	0.498501	0.387365	1000
9p21.3	*MTAP*	3	142	0.511489	0.385122	1000
9p21.3	*DMRTA1*	6	99	0.458541	0.361992	1000
9q31.3	*ACTL7B*	7	225	0.0243951	0.357704	5000
9q31.3	*ACTL7A*	3	15	0.877123	0.39043	1000
9q31.3	*IKBKAP*	25	532	0.277722	0.439947	1000
9q31.3	*CTNNAL1*	3	148	0.917083	0.412994	1000
9q31.3	*TMEM245*	2	2	0.707293	0.415635	1000
10q24.32	*TMEM180*	4	271	0.793207	0.372718	1000
10q24.32	*SUFU*	4	28	0.953047	0.381939	1000
10q24.32	*TRIM8*	2	2	0.696304	0.428279	1000
10q24.32	*WBP1L*	6	164	0.863137	0.344598	1000
10q24.32	*CYP17A1*	4	208	0.84016	0.36976	1000

10q24.32	*AS3MT*	2	24	0.874126	0.409238	1000
10q24.32	*C10orf32-AS3MT*	2	24	0.866134	0.400477	1000
10q24.32	*CNNM2*	4	161	0.0875825	0.358167	5000
10q24.32	*NT5C2*	3	154	0.704296	0.383839	1000
10q24.32	*INA*	5	27	0.00279944	0.455156	5000
10q24.32	*TAF5*	3	85	0.833167	0.381828	1000
10q24.32	*PDCD11*	16	322	0.306693	0.356928	1000
10q24.32	*CALHM2*	6	18	0.04979	0.434441	5000
10q24.32	*CALHM1*	6	241	0.537463	0.433275	1000
13q13.1	*PDS5B*	3	117	0.47952	0.342009	1000
13q13.1	*KL*	12	278	0.401598	0.388544	1000
13q13.1	*STARD13*	10	360	**0.00079984**	0.351073	5000
18q11.2	*GATA6*	2	2	0.731269	0.441789	1000
18q11.2	*CTAGE1*	6	174	0.0753849	0.371797	5000
18q11.2	*RBBP8*	6	43	0.0585883	0.376427	5000
18q11.2	*CABLES1*	3	68	0.00559888	0.356104	5000

BIBLIOGRAPHY

1. Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. Nature 488, 370-374.

2. Bjerregaard, P., Young, T.K., Dewailly, E., and Ebbesson, S.O. (2004). Indigenous health in the Arctic: an overview of the circumpolar Inuit population. Scand J Public Health 32, 390-395.

3. Young, T.K., and Bjerregaard, P. (2008). Health transitions in Arctic populations.(Toronto ; Buffalo: University of Toronto Press).

4. Prowse, T.D., and Furgal, C. (2009). Northern Canada in a changing climate: major findings and conclusions. Ambio 38, 290-292.

5. Tvermosegaard, M., Dahl-Petersen, I.K., Nielsen, N.O., Bjerregaard, P., and Jorgensen, M.E. (2015). Cardiovascular Disease Susceptibility and Resistance in Circumpolar Inuit Populations. Can J Cardiol 31, 1116-1123.

6. Lee, R.S., Radomski, N., Proulx, J.F., Levade, I., Shapiro, B.J., McIntosh, F., Soualhine, H., Menzies, D., and Behr, M.A. (2015). Population genomics of Mycobacterium tuberculosis in the Inuit. Proc Natl Acad Sci U S A 112, 13609-13614.

7. Friborg, J.T., and Melbye, M. (2008). Cancer patterns in Inuit populations. Lancet Oncol 9, 892-900.

8. Campbell, J.M., and Arctic Institute of North America. (1962). Prehistoric cultural relations between the arctic and temperate zones of North America.(Montreal: Arctic Institute of North America).

9. Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T.S., Gronnow, B., Appelt, M., Gullov, H.C., Friesen, T.M., et al. (2014). The genetic prehistory of the New World Arctic. Science 345, 1255832.

10. Sturtevant, W.C. (1978). Handbook of North American Indians.(Washington: Smithsonian Institution : For sale by the Supt. of Docs., U.S. G.P.O.).

11. Arneborg, J., Grønnow, B., Greenland Research Centre (Nationalmuseet : Denmark), North Atlantic Biocultural Organisation., and Nationalmuseet (Denmark). (2006). Dynamics of northern societies : proceedings of the SILA/NABO Conference on Arctic and North Atlantic Archaeology, Copenhagen, May 10th-14th, 2004.(Copenhagen: National Museum of Denmark).

12. Crowe, K.J. (1991). A history of the original peoples of northern Canada.(Montreal ; Buffalo: McGill-Queen's University Press).

13. McCullough, K.M., Canadian Museum of Civilization., and Archaeological Survey of Canada. (1989). The Ruin Islanders : Thule culture pioneers in the Eastern High Arctic.(Hull, Quebec: Canadian Museum of Civilization).

14. Moltke, I., Fumagalli, M., Korneliussen, T.S., Crawford, J.E., Bjerregaard, P., Jorgensen, M.E., Grarup, N., Gullov, H.C., Linneberg, A., Pedersen, O., et al. (2015). Uncovering the genetic history of the present-day Greenlandic population. Am J Hum Genet 96, 54-69.

15. Helgason, A., Palsson, G., Pedersen, H.S., Angulalik, E., Gunnarsdottir, E.D., Yngvadottir, B., and Stefansson, K. (2006). mtDNA variation in Inuit populations of Greenland and Canada: migration history and population structure. Am J Phys Anthropol 130, 123-134.

16. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463, 757-762.

17. Olofsson, J.K., Pereira, V., Borsting, C., and Morling, N. (2015). Peopling of the North Circumpolar Region--insights from Y chromosome STR and SNP typing of Greenlanders. PLoS One 10, e0116573.

18. H.C, G. (2008). The Nature of Contact between Native Greenlanders and Norse. Journal of the North Atlantic, 16–24.

19. J, B. (1986). The decline of the Norse settlements in Greenland. Arctic Anthropology 23, 109-135.

20. M Appelt, H.G. (2009). Tunit, Norsemen, and Inuit in Thirteenth-Century Northwest Greenland: Dorset between the Devil and the Deep Sea. The Northern World AD, 900-1400.

21. Bosch, E., Calafell, F., Rosser, Z.H., Norby, S., Lynnerup, N., Hurles, M.E., and Jobling, M.A. (2003). High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. Hum Genet 112, 353-363.

22. Dyke, J.M.S.a.A.S. (2002). Variability in Palaeoeskimo Occupation on South-Western Victoria Island, Arctic Canada: Causes and Consequences. World Archaeology 33, 508-522.

23. Gilbert, M.T., Kivisild, T., Gronnow, B., Andersen, P.K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Gotherstrom, A., Campos, P.F., et al. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science 320, 1787-1789.

24. Raff, J.A., Rzhetskaya, M., Tackney, J., and Hayes, M.G. (2015). Mitochondrial diversity of Inupiat people from the Alaskan North Slope provides evidence for the origins of the Paleo- and Neo-Eskimo peoples. Am J Phys Anthropol 157, 603-614.

25. Dryomov, S.V., Nazhmidenova, A.M., Shalaurova, S.A., Morozov, I.V., Tabarev, A.V., Starikovskaya, E.B., and Sukernik, R.I. (2015). Mitochondrial genome diversity at the Bering Strait area highlights prehistoric human migrations from Siberia to northern North America. Eur J Hum Genet 23, 1399-1404.

26. Dulik, M.C., Owings, A.C., Gaieski, J.B., Vilar, M.G., Andre, A., Lennie, C., Mackenzie, M.A., Kritsch, I., Snowshoe, S., Wright, R., et al. (2012). Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. Proc Natl Acad Sci U S A 109, 8471-8476.

27. Pereira, V., Tomas, C., Sanchez, J.J., Syndercombe-Court, D., Amorim, A., Gusmao, L., Prata, M.J., and Morling, N. (2015). The peopling of Greenland: further insights from the analysis of genetic diversity using autosomal and X-chromosomal markers. Eur J Hum Genet 23, 245-251.

28. Tomas, C., Mogensen, H.S., Friis, S.L., Hallenberg, C., Stene, M.C., and Morling, N. (2014). Concordance study and population frequencies for 16 autosomal STRs analyzed with PowerPlex(R) ESI 17 and AmpFlSTR(R) NGM SElect in Somalis, Danes and Greenlanders. Forensic Sci Int Genet 11, e18-21.

29. Fumagalli, M., Moltke, I., Grarup, N., Racimo, F., Bjerregaard, P., Jorgensen, M.E., Korneliussen, T.S., Gerbault, P., Skotte, L., Linneberg, A., et al. (2015). Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science 349, 1343-1347.

30. Moltke, I., Grarup, N., Jorgensen, M.E., Bjerregaard, P., Treebak, J.T., Fumagalli, M., Korneliussen, T.S., Andersen, M.A., Nielsen, T.S., Krarup, N.T., et al. (2014). A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature 512, 190-193.

31. Snodgrass, J.J., Leonard, W.R., Sorensen, M.V., Tarskaia, L.A., and Mosher, M.J. (2008). The influence of basal metabolic rate on blood pressure among indigenous Siberians. Am J Phys Anthropol 137, 145-155.

32. Snodgrass, J.J., Leonard, W.R., Tarskaia, L.A., Alekseev, V.P., and Krivoshapkin, V.G. (2005). Basal metabolic rate in the Yakut (Sakha) of Siberia. Am J Hum Biol 17, 155-172.

33. Starikovskaya, Y.B., Sukernik, R.I., Schurr, T.G., Kogelnik, A.M., and Wallace, D.C. (1998). mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World. Am J Hum Genet 63, 1473-1491.

34. Volodko, N.V., Starikovskaya, E.B., Mazunin, I.O., Eltsov, N.P., Naidenko, P.V., Wallace, D.C., and Sukernik, R.I. (2008). Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. Am J Hum Genet 82, 1084-1100.

35. Derbeneva, O.A., Sukernik, R.I., Volodko, N.V., Hosseini, S.H., Lott, M.T., and Wallace, D.C. (2002). Analysis of mitochondrial DNA diversity in the aleuts of the commander islands and its implications for the genetic history of beringia. Am J Hum Genet 71, 415-421.

36. Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.J., and Norby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67, 718-726.

37. Hancock, A.M., Witonsky, D.B., Alkorta-Aranburu, G., Beall, C.M., Gebremedhin, A., Sukernik, R., Utermann, G., Pritchard, J.K., Coop, G., and Di Rienzo, A. (2011). Adaptations to climate-mediated selective pressures in humans. PLoS Genet 7, e1001375.

38. Hancock, A.M., Witonsky, D.B., Gordon, A.S., Eshel, G., Pritchard, J.K., Coop, G., and Di Rienzo, A. (2008). Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet 4, e32.

39. Cardona, A., Pagani, L., Antao, T., Lawson, D.J., Eichstaedt, C.A., Yngvadottir, B., Shwe, M.T., Wee, J., Romero, I.G., Raj, S., et al. (2014). Genome-wide analysis of cold adaptation in indigenous Siberian populations. PLoS One 9, e98076.

40. Clemente, F.J., Cardona, A., Inchley, C.E., Peter, B.M., Jacobs, G., Pagani, L., Lawson, D.J., Antao, T., Vicente, M., Mitt, M., et al. (2014). A Selective Sweep on a Deleterious Mutation in CPT1A in Arctic Populations. Am J Hum Genet 95, 584-589.

41. Li, S.W., Wang, J., Yang, Y., Liu, Z.J., Cheng, L., Liu, H.Y., Ma, P., Luo, W., and Liu, S.M. (2016). Polymorphisms in FADS1 and FADS2 alter plasma fatty acids and desaturase levels in type 2 diabetic patients with coronary artery disease. J Transl Med 14, 79.

42. Yang, Q., Yin, R.X., Cao, X.L., Wu, D.F., Chen, W.X., and Zhou, Y.J. (2015). Association of two polymorphisms in the FADS1/FADS2 gene cluster and the risk of coronary artery disease and ischemic stroke. Int J Clin Exp Pathol 8, 7318-7331.

43. Voruganti, V.S., Higgins, P.B., Ebbesson, S.O., Kennish, J., Goring, H.H., Haack, K., Laston, S., Drigalenko, E., Wenger, C.R., Harris, W.S., et al. (2012). Variants in CPT1A, FADS1, and FADS2 are Associated with Higher Levels of Estimated Plasma and Erythrocyte Delta-5 Desaturases in Alaskan Eskimos. Front Genet 3, 86.

44. Eaton, S.B., Eaton, S.B., 3rd, Sinclair, A.J., Cordain, L., and Mann, N.J. (1998). Dietary intake of long-chain polyunsaturated fatty acids during the paleolithic. World Rev Nutr Diet 83, 12-23.

45. Ameur, A., Enroth, S., Johansson, A., Zaboli, G., Igl, W., Johansson, A.C., Rivas, M.A., Daly, M.J., Schmitz, G., Hicks, A.A., et al. (2012). Genetic adaptation of fatty-acid metabolism: a human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. Am J Hum Genet 90, 809-820.

46. Mathias, R.A., Fu, W., Akey, J.M., Ainsworth, H.C., Torgerson, D.G., Ruczinski, I., Sergeant, S., Barnes, K.C., and Chilton, F.H. (2012). Adaptive evolution of the FADS gene cluster within Africa. PLoS One 7, e44926.

47. Kumar S.D. Kothapalli1, K.Y., Maithili S. Gadgil, Susan E. Carlson, Kimberly O. O'Brien, Ji Yao Zhang, Hui Gyu Park, Kinsley Ojukwu, James Zou, Stephanie S. Hyon, Kalpana S. Joshi, Zhenglong Gu, Alon Keinan, and J. Thomas Brenna. (2016). Positive selection on a regulatory insertion-deletion polymorphism in FADS2 influences apparent endogenous synthesis of arachidonic acid. Molecular Biology and Evolution.

48. Collins, S.A., Sinclair, G., McIntosh, S., Bamforth, F., Thompson, R., Sobol, I., Osborne, G., Corriveau, A., Santos, M., Hanley, B., et al. (2010). Carnitine palmitoyltransferase 1A (CPT1A) P479L prevalence in live newborns in Yukon, Northwest Territories, and Nunavut. Mol Genet Metab 101, 200-204.

49. Zhou, S., Xiong, L., Xie, P., Ambalavanan, A., Bourassa, C.V., Dionne-Laporte, A., Spiegelman, D., Turcotte Gauthier, M., Henrion, E., Diallo, O., et al. (2015). Increased missense mutation burden of Fatty Acid metabolism related genes in nunavik inuit population. PLoS One 10, e0128255.

50. Gillingham, M.B., Hirschfeld, M., Lowe, S., Matern, D., Shoemaker, J., Lambert, W.E., and Koeller, D.M. (2011). Impaired fasting tolerance among Alaska native children with a common carnitine palmitoyltransferase 1A sequence variant. Mol Genet Metab 104, 261-264.

51. Collins, S.A., Surmala, P., Osborne, G., Greenberg, C., Bathory, L.W., Edmunds-Potvin, S., and Arbour, L. (2012). Causes and risk factors for infant mortality in Nunavut, Canada 1999-2011. BMC Pediatr 12, 190.

52. Brown, N.F., Mullur, R.S., Subramanian, I., Esser, V., Bennett, M.J., Saudubray, J.M., Feigenbaum, A.S., Kobari, J.A., Macleod, P.M., McGarry, J.D., et al. (2001). Molecular characterization of L-CPT I deficiency in six patients: insights into function of the native enzyme. J Lipid Res 42, 1134-1142.

53. Greenberg, C.R., Dilling, L.A., Thompson, G.R., Seargeant, L.E., Haworth, J.C., Phillips, S., Chan, A., Vallance, H.D., Waters, P.J., Sinclair, G., et al. (2009). The paradox of the carnitine palmitoyltransferase type Ia P479L variant in Canadian Aboriginal populations. Mol Genet Metab 96, 201-207.

54. McVean, G. (2007). The structure of linkage disequilibrium around a selective sweep. Genetics 175, 1395-1406.

55. Khrameeva, E.E., Bozek, K., He, L., Yan, Z., Jiang, X., Wei, Y., Tang, K., Gelfand, M.S., Prufer, K., Kelso, J., et al. (2014). Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. Nat Commun 5, 3584.

56. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505, 87-91.

57. Jorgensen, M.E., Borch-Johnsen, K., and Bjerregaard, P. (2006). Lifestyle modifies obesity-associated risk of cardiovascular disease in a genetically homogeneous population. Am J Clin Nutr 84, 29-36.

58. Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D.G., Mulligan, C.J., Bravi, C.M., Rickards, O., Martinez-Labarga, C., Khusnutdinova, E.K., et al. (2007). Beringian standstill and spread of Native American founders. PLoS One 2, e829.

59. Johansen, C.T., Gallinger, Z.R., Wang, J., Ban, M.R., Young, T.K., Bjerregaard, P., and Hegele, R.A. (2010). Rare ATGL haplotypes are associated with increased plasma triglyceride concentrations in the Greenland Inuit. Int J Circumpolar Health 69, 3-12.

60. Bjerregaard, P., Mikkelsen, S.S., Becker, U., Hansen, T., and Tolstrup, J.S. (2014). Genetic variation in alcohol metabolizing enzymes among Inuit and its relation to drinking patterns. Drug Alcohol Depend 144, 239-244.

61. Ghisari, M., Long, M., and Bonefeld-Jorgensen, E.C. (2013). Genetic polymorphisms in CYP1A1, CYP1B1 and COMT genes in Greenlandic Inuit and Europeans. Int J Circumpolar Health 72, 21113.

62. Sharp, D. (2009). Environmental toxins, a potential risk factor for diabetes among Canadian Aboriginals. Int J Circumpolar Health 68, 316-326.

63. Rudkowska, I., Dewailly, E., Hegele, R.A., Boiteau, V., Dube-Linteau, A., Abdous, B., Giguere, Y., Chateau-Degat, M.L., and Vohl, M.C. (2013). Gene-diet interactions on plasma lipid levels in the Inuit population. Br J Nutr 109, 953-961.

64. Vaughan, L.K., Wiener, H.W., Aslibekyan, S., Allison, D.B., Havel, P.J., Stanhope, K.L., O'Brien, D.M., Hopkins, S.E., Lemas, D.J., Boyer, B.B., et al. (2015). Linkage and association analysis of obesity traits reveals novel loci and interactions with dietary n-3 fatty acids in an Alaska Native (Yup'ik) population. Metabolism 64, 689-697.

65. Voruganti, V.S., Cole, S.A., Ebbesson, S.O., Goring, H.H., Haack, K., Laston, S., Wenger, C.R., Tejero, M.E., Devereux, R.B., Fabsitz, R.R., et al. (2010). Genetic variation in APOJ, LPL, and TNFRSF10B affects plasma fatty acid distribution in Alaskan Eskimos. Am J Clin Nutr 91, 1574-1583.

66. Klimentidis, Y.C., Lemas, D.J., Wiener, H.H., O'Brien, D.M., Havel, P.J., Stanhope, K.L., Hopkins, S.E., Tiwari, H.K., and Boyer, B.B. (2014). CDKAL1 and HHEX are associated with type 2 diabetes-related traits among Yup'ik people. J Diabetes 6, 251-259.

67. Dube, J.B., Wang, J., Cao, H., McIntyre, A.D., Johansen, C.T., Hopkins, S.E., Stringer, R., Hosseinzadeh, S., Kennedy, B.A., Ban, M.R., et al. (2015). Common low-density lipoprotein

receptor p.G116S variant has a large effect on plasma low-density lipoprotein cholesterol in circumpolar inuit populations. Circ Cardiovasc Genet 8, 100-105.

68. Rousseau-Nepton, I., Okubo, M., Grabs, R., Consortium, F.C., Mitchell, J., Polychronakos, C., and Rodd, C. (2015). A founder AGL mutation causing glycogen storage disease type IIIa in Inuit identified through whole-exome sequencing: a case series. CMAJ 187, E68-73.

69. Marcadier, J.L., Boland, M., Scott, C.R., Issa, K., Wu, Z., McIntyre, A.D., Hegele, R.A., Geraghty, M.T., and Lines, M.A. (2015). Congenital sucrase-isomaltase deficiency: identification of a common Inuit founder mutation. CMAJ 187, 102-107.

70. Gauthier, B., Coutlee, F., Franco, E.L., and Brassard, P. (2015). Human papillomavirus variants among Inuit women in northern Quebec, Canada. Int J Circumpolar Health 74, 29482.

71. Metcalfe, S., Roger, M., Faucher, M.C., Coutlee, F., Franco, E.L., and Brassard, P. (2013). The association between human leukocyte antigen (HLA)-G polymorphisms and human papillomavirus (HPV) infection in Inuit women of northern Quebec. Hum Immunol 74, 1610-1615.

72. Khoo, S.K., Zhang, G., Backer, V., Porsbjerg, C., Nepper-Christensen, S., Creegan, R., Baynam, G., de Klerk, N., Rossi, G.A., Hagel, I., et al. (2006). Associations of a novel IL4RA polymorphism, Ala57Thr, in Greenlander Inuit. J Allergy Clin Immunol 118, 627-634.

73. Candelaria, P.V., Backer, V., Khoo, S.K., Bizzintino, J.A., Hayden, C.M., Baynam, G., Laing, I.A., Zhang, G., Porsbjerg, C., Goldblatt, J., et al. (2010). The importance of environment on respiratory genotype/phenotype relationships in the Inuit. Allergy 65, 229-237.

74. Li, L., Hamel, N., Baker, K., McGuffin, M.J., Couillard, M., Gologan, A., Marcus, V.A., Chodirker, B., Chudley, A., Stefanovici, C., et al. (2015). A homozygous PMS2 founder mutation with an attenuated constitutional mismatch repair deficiency phenotype. J Med Genet 52, 348-352.

75. Hansen, T.V., Ejlertsen, B., Albrechtsen, A., Bergsten, E., Bjerregaard, P., Hansen, T., Myrhoj, T., Nielsen, P.B., Timmermans-Wielenga, V., Andersen, M.K., et al. (2009). A common Greenlandic Inuit BRCA1 RING domain founder mutation. Breast Cancer Res Treat 115, 69-76.

76. Harboe, T.L., Eiberg, H., Kern, P., Ejlertsen, B., Nedergaard, L., Timmermans-Wielenga, V., Nielsen, I.M., and Bisgaard, M.L. (2009). A high frequent BRCA1 founder mutation identified in the Greenlandic population. Fam Cancer 8, 413-419.

77. Ghisari, M., Eiberg, H., Long, M., and Bonefeld-Jorgensen, E.C. (2014). Polymorphisms in phase I and phase II genes and breast cancer risk and relations to persistent organic pollutant exposure: a case-control study in Inuit women. Environ Health 13, 19.

78. Giwercman, C., Giwercman, A., Pedersen, H.S., Toft, G., Lundin, K., Bonde, J.P., and Lundberg Giwercman, Y. (2008). Polymorphisms in genes regulating androgen activity among prostate cancer low-risk Inuit men and high-risk Scandinavians. Int J Androl 31, 25-30.

79. Nielsen, I.M., and Eiberg, H. (2004). Cholestasis Familiaris Groenlandica: an epidemiological, clinical and genetic study. Int J Circumpolar Health 63 Suppl 2, 192-194.

80. Andersen, S., Okkels, H., Krarup, H., and Laurberg, P. (2006). Geographical clustering and maintained health in individuals harbouring the mutation for Greenland familial cholestasis: A population-based study. Scand J Gastroenterol 41, 445-450.

81. Nytofte, N.S., Serrano, M.A., Monte, M.J., Gonzalez-Sanchez, E., Tumer, Z., Ladefoged, K., Briz, O., and Marin, J.J. (2011). A homozygous nonsense mutation (c.214C->A) in the biliverdin reductase alpha gene (BLVRA) results in accumulation of biliverdin during episodes of cholestasis. J Med Genet 48, 219-225.

82. Eiberg, H., Hansen, L., Korbo, L., Nielsen, I.M., Svenstrup, K., Bech, S., Pinborg, L.H., Friberg, L., Hjermind, L.E., Olsen, O.R., et al. (2012). Novel mutation in ATP13A2 widens the spectrum of Kufor-Rakeb syndrome (PARK9). Clin Genet 82, 256-263.

83. Homoe, P., Koch, A., Rendtorff, N.D., Lodahl, M., Andersen, T., Andersen, S., Eiberg, H., Nielsen, I.M., and Tranebjaerg, L. (2012). GJB2 (Connexin-26) mutations are not frequent among hearing impaired patients in east Greenland. Int J Audiol 51, 433-436.

84. Chateau-Degat, M.L., Dewailly, E., Louchini, R., Counil, E., Noel, M., Ferland, A., Lucas, M., Valera, B., Ekoe, J.M., Ladouceur, R., et al. (2010). Cardiovascular burden and related risk factors among Nunavik (Quebec) Inuit: insights from baseline findings in the circumpolar Inuit health in transition cohort study. Can J Cardiol 26, 190-196.

85. Day, G.E., and Lanier, A.P. (2003). Alaska native mortality, 1979-1998. Public Health Rep 118, 518-530.

86. Lindgaard, L., Eskesen, V., Gjerris, F., and Olsen, N.V. (2003). Familial aggregation of intracranial aneurysms in an Inuit patient population in Kalaallit Nunaat (Greenland). Neurosurgery 52, 357-362; discussion 362-353.

87. Ostergaard Kristensen, M. (1983). Increased incidence of bleeding intracranial aneurysms in Greenlandic Eskimos. Acta Neurochir (Wien) 67, 37-43.

88. Ronkainen, A., Hernesniemi, J., and Ryynanen, M. (1993). Familial subarachnoid hemorrhage in east Finland, 1977-1990. Neurosurgery 33, 787-796; discussion 796-797.

89. Kurki, M.I., Gaal, E.I., Kettunen, J., Lappalainen, T., Menelaou, A., Anttila, V., van 't Hof, F.N., von Und Zu Fraunberg, M., Helisalmi, S., Hiltunen, M., et al. (2014). High risk population isolate reveals low frequency variants predisposing to intracranial aneurysms. PLoS Genet 10, e1004134.

90. Mortality, G.B.D., and Causes of Death, C. (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 385, 117-171.

91. Donnan, G.A., Fisher, M., Macleod, M., and Davis, S.M. (2008). Stroke. Lancet 371, 1612-1623.

92. Feigin, V.L., Rinkel, G.J., Lawes, C.M., Algra, A., Bennett, D.A., van Gijn, J., and Anderson, C.S. (2005). Risk factors for subarachnoid hemorrhage: an updated systematic review of epidemiological studies. Stroke 36, 2773-2780.

93. Stegmayr, B., Eriksson, M., and Asplund, K. (2004). Declining mortality from subarachnoid hemorrhage: changes in incidence and case fatality from 1985 through 2000. Stroke 35, 2059-2063.

94. de Rooij, N.K., Linn, F.H., van der Plas, J.A., Algra, A., and Rinkel, G.J. (2007). Incidence of subarachnoid haemorrhage: a systematic review with emphasis on region, age, gender and time trends. J Neurol Neurosurg Psychiatry 78, 1365-1372.

95. Linn, F.H., Rinkel, G.J., Algra, A., and van Gijn, J. (1996). Incidence of subarachnoid hemorrhage: role of region, year, and rate of computed tomography: a meta-analysis. Stroke 27, 625-629.

96. Vlak, M.H., Algra, A., Brandenburg, R., and Rinkel, G.J. (2011). Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: a systematic review and meta-analysis. Lancet Neurol 10, 626-636.

97. Stehbens, W. (1983). The pathology of intracranial arterial aneurysms and their complications. In Intracranial Aneurysms J.L. Fox, ed. (Springer

98. Rinkel, G.J., Djibuti, M., Algra, A., and van Gijn, J. (1998). Prevalence and risk of rupture of intracranial aneurysms: a systematic review. Stroke 29, 251-256.

99. Brown, R.D., Jr., and Broderick, J.P. (2014). Unruptured intracranial aneurysms: epidemiology, natural history, management options, and familial screening. Lancet Neurol 13, 393-404.

100. Tromp, G., Weinsheimer, S., Ronkainen, A., and Kuivaniemi, H. (2014). Molecular basis and genetic predisposition to intracranial aneurysm. Ann Med 46, 597-606.

101. Morita, A., Kimura, T., Shojima, M., Sameshima, T., and Nishihara, T. (2010). Unruptured intracranial aneurysms: current perspectives on the origin and natural course, and quest for standards in the management strategy. Neurol Med Chir (Tokyo) 50, 777-787.

102. Vernooij, M.W., Ikram, M.A., Tanghe, H.L., Vincent, A.J., Hofman, A., Krestin, G.P., Niessen, W.J., Breteler, M.M., and van der Lugt, A. (2007). Incidental findings on brain MRI in the general population. N Engl J Med 357, 1821-1828.

103. Korja, M., and Kaprio, J. (2016). Controversies in epidemiology of intracranial aneurysms and SAH. Nat Rev Neurol 12, 50-55.

104. Schievink, W.I., Riedinger, M., and Maya, M.M. (2005). Frequency of incidental intracranial aneurysms in neurofibromatosis type 1. Am J Med Genet A 134A, 45-48.

105. Rubinstein, M.K., and Cohen, N.H. (1964). Ehlers-Danlos Syndrome Associated with Multiple Intracranial Aneurysms. Neurology 14, 125-132.

106. Schievink, W.I. (1999). Marfan syndrome and intracranial aneurysms. Stroke 30, 2767-2768.

107. Kim, H., Nelson, J., Krings, T., terBrugge, K.G., McCulloch, C.E., Lawton, M.T., Young, W.L., Faughnan, M.E., and Brain Vascular Malformation Consortium, H.H.T.I.G. (2015). Hemorrhage rates from brain arteriovenous malformation in patients with hereditary hemorrhagic telangiectasia. Stroke 46, 1362-1364.

108. Adachi, K., Kudo, M., Chen, M.N., Nakazawa, S., and Wakabayashi, I. (1993). Cerebral aneurysm associated with multiple endocrine neoplasia, type 1--case report. Neurol Med Chir (Tokyo) 33, 309-311.

109. Wiebers, D.O., Whisnant, J.P., Huston, J., 3rd, Meissner, I., Brown, R.D., Jr., Piepgras, D.G., Forbes, G.S., Thielen, K., Nichols, D., O'Fallon, W.M., et al. (2003). Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment. Lancet 362, 103-110.

110. Korja, M., Lehto, H., and Juvela, S. (2014). Lifelong rupture risk of intracranial aneurysms depends on risk factors: a prospective Finnish cohort study. Stroke 45, 1958-1963.

111. Greving, J.P., Wermer, M.J., Brown, R.D., Jr., Morita, A., Juvela, S., Yonekura, M., Ishibashi, T., Torner, J.C., Nakayama, T., Rinkel, G.J., et al. (2014). Development of the PHASES score for prediction of risk of rupture of intracranial aneurysms: a pooled analysis of six prospective cohort studies. Lancet Neurol 13, 59-66.

112. Molyneux, A.J., Kerr, R.S., Yu, L.M., Clarke, M., Sneade, M., Yarnold, J.A., Sandercock, P., and International Subarachnoid Aneurysm Trial Collaborative, G. (2005). International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and aneurysm occlusion. Lancet 366, 809-817.

113. Korja, M., Silventoinen, K., Laatikainen, T., Jousilahti, P., Salomaa, V., Hernesniemi, J., and Kaprio, J. (2013). Risk factors and their combined effects on the incidence rate of subarachnoid hemorrhage--a population-based cohort study. PLoS One 8, e73760.

114. Qureshi, A.I., Suri, M.F., Yahia, A.M., Suarez, J.I., Guterman, L.R., Hopkins, L.N., and Tamargo, R.J. (2001). Risk factors for subarachnoid hemorrhage. Neurosurgery 49, 607-612; discussion 612-603.

115. Turjman, A.S., Turjman, F., and Edelman, E.R. (2014). Role of fluid dynamics and inflammation in intracranial aneurysm formation. Circulation 129, 373-382.

116. Munarriz, P.M., Gomez, P.A., Paredes, I., Castano-Leon, A.M., Cepeda, S., and Lagares, A. (2016). Basic Principles of Hemodynamics and Cerebral Aneurysms. World Neurosurg 88, 311-319.

117. Chalouhi, N., Ali, M.S., Jabbour, P.M., Tjoumakaris, S.I., Gonzalez, L.F., Rosenwasser, R.H., Koch, W.J., and Dumont, A.S. (2012). Biology of intracranial aneurysms: role of inflammation. J Cereb Blood Flow Metab 32, 1659-1676.

118. Hashimoto, T., Meng, H., and Young, W.L. (2006). Intracranial aneurysms: links among inflammation, hemodynamics and vascular remodeling. Neurol Res 28, 372-380.

119. Finlay, H.M., Whittaker, P., and Canham, P.B. (1998). Collagen organization in the branching region of human brain arteries. Stroke 29, 1595-1601.

120. Stehbens, W.E. (1989). Etiology of intracranial berry aneurysms. J Neurosurg 70, 823-831.

121. Kataoka, K., Taneda, M., Asai, T., Kinoshita, A., Ito, M., and Kuroda, R. (1999). Structural fragility and inflammatory response of ruptured cerebral aneurysms. A comparative study between ruptured and unruptured cerebral aneurysms. Stroke 30, 1396-1401.

122. Guo, F., Li, Z., Song, L., Han, T., Feng, Q., Guo, Y., Xu, J., He, M., and You, C. (2007). Increased apoptosis and cysteinyl aspartate specific protease-3 gene expression in human intracranial aneurysm. J Clin Neurosci 14, 550-555.

123. Frosen, J., Piippo, A., Paetau, A., Kangasniemi, M., Niemela, M., Hernesniemi, J., and Jaaskelainen, J. (2004). Remodeling of saccular cerebral artery aneurysm wall is associated with rupture: histological analysis of 24 unruptured and 42 ruptured cases. Stroke 35, 2287-2293.

124. Kilic, T., Sohrabifar, M., Kurtkaya, O., Yildirim, O., Elmaci, I., Gunel, M., and Pamir, M.N. (2005). Expression of structural proteins and angiogenic factors in normal arterial and unruptured and ruptured aneurysm walls. Neurosurgery 57, 997-1007; discussion 1997-1007.

125. Austin, G., Fisher, S., Dickson, D., Anderson, D., and Richardson, S. (1993). The significance of the extracellular matrix in intracranial aneurysms. Ann Clin Lab Sci 23, 97-105.

126. Kosierkiewicz, T.A., Factor, S.M., and Dickson, D.W. (1994). Immunocytochemical studies of atherosclerotic lesions of cerebral berry aneurysms. J Neuropathol Exp Neurol 53, 399-406.

127. Alfano, J.M., Kolega, J., Natarajan, S.K., Xiang, J., Paluch, R.A., Levy, E.I., Siddiqui, A.H., and Meng, H. (2013). Intracranial aneurysms occur more frequently at bifurcation sites that typically experience higher hemodynamic stresses. Neurosurgery 73, 497-505.

128. Meng, H., Wang, Z., Hoi, Y., Gao, L., Metaxa, E., Swartz, D.D., and Kolega, J. (2007). Complex hemodynamics at the apex of an arterial bifurcation induces vascular remodeling resembling cerebral aneurysm initiation. Stroke 38, 1924-1931.

129. Metaxa, E., Tremmel, M., Natarajan, S.K., Xiang, J., Paluch, R.A., Mandelbaum, M., Siddiqui, A.H., Kolega, J., Mocco, J., and Meng, H. (2010). Characterization of critical hemodynamics contributing to aneurysmal remodeling at the basilar terminus in a rabbit model. Stroke 41, 1774-1782.

130. Sakamoto, N., Saito, N., Han, X., Ohashi, T., and Sato, M. (2010). Effect of spatial gradient in fluid shear stress on morphological changes in endothelial cells in response to flow. Biochem Biophys Res Commun 395, 264-269.

131. Szymanski, M.P., Metaxa, E., Meng, H., and Kolega, J. (2008). Endothelial cell layer subjected to impinging flow mimicking the apex of an arterial bifurcation. Ann Biomed Eng 36, 1681-1689.

132. Aoki, T., Nishimura, M., Matsuoka, T., Yamamoto, K., Furuyashiki, T., Kataoka, H., Kitaoka, S., Ishibashi, R., Ishibazawa, A., Miyamoto, S., et al. (2011). PGE(2) -EP(2) signalling in endothelium is activated by haemodynamic stress and induces cerebral aneurysm through an amplifying loop via NF-kappaB. Br J Pharmacol 163, 1237-1249.

133. Resnick, N., Yahav, H., Shay-Salit, A., Shushy, M., Schubert, S., Zilberman, L.C., and Wofovitz, E. (2003). Fluid shear stress and the vascular endothelium: for better and for worse. Prog Biophys Mol Biol 81, 177-199.

134. Ando, J., Ohtsuka, A., Korenaga, R., Kawamura, T., and Kamiya, A. (1993). Wall shear stress rather than shear rate regulates cytoplasmic Ca++ responses to flow in vascular endothelial cells. Biochem Biophys Res Commun 190, 716-723.

135. Resnick, N., Yahav, H., Khachigian, L.M., Collins, T., Anderson, K.R., Dewey, F.C., and Gimbrone, M.A., Jr. (1997). Endothelial gene regulation by laminar shear stress. Adv Exp Med Biol 430, 155-164.

136. Helmke, B.P., and Davies, P.F. (2002). The cytoskeleton under external fluid mechanical forces: hemodynamic forces acting on the endothelium. Ann Biomed Eng 30, 284-296.

137. Malek, A.M., and Izumo, S. (1996). Mechanism of endothelial cell shape change and cytoskeletal remodeling in response to fluid shear stress. J Cell Sci 109 (Pt 4), 713-726.

138. Nakajima, N., Nagahiro, S., Sano, T., Satomi, J., and Satoh, K. (2000). Phenotypic modulation of smooth muscle cells in human cerebral aneurysmal walls. Acta Neuropathol 100, 475-480.

139. Gibbons, G.H., and Dzau, V.J. (1994). The emerging concept of vascular remodeling. N Engl J Med 330, 1431-1438.

140. Malek, A.M., Alper, S.L., and Izumo, S. (1999). Hemodynamic shear stress and its role in atherosclerosis. JAMA 282, 2035-2042.

141. Sho, E., Komatsu, M., Sho, M., Nanjo, H., Singh, T.M., Xu, C., Masuda, H., and Zarins, C.K. (2003). High flow drives vascular endothelial cell proliferation during flow-induced arterial remodeling associated with the expression of vascular endothelial growth factor. Exp Mol Pathol 75, 1-11.

142. Dolan, J.M., Kolega, J., and Meng, H. (2013). High wall shear stress and spatial gradients in vascular pathology: a review. Ann Biomed Eng 41, 1411-1427.

143. Yoshida, T., and Owens, G.K. (2005). Molecular determinants of vascular smooth muscle cell diversity. Circ Res 96, 280-291.

144. Tulamo, R., Frosen, J., Junnikkala, S., Paetau, A., Kangasniemi, M., Pelaez, J., Hernesniemi, J., Niemela, M., and Meri, S. (2010). Complement system becomes activated by the classical pathway in intracranial aneurysm walls. Lab Invest 90, 168-179.

145. Frosen, J., Tulamo, R., Heikura, T., Sammalkorpi, S., Niemela, M., Hernesniemi, J., Levonen, A.L., Horkko, S., and Yla-Herttuala, S. (2013). Lipid accumulation, lipid oxidation, and low plasma levels of acquired antibodies against oxidized lipids associate with degeneration and rupture of the intracranial aneurysm wall. Acta Neuropathol Commun 1, 71.

146. Kanematsu, Y., Kanematsu, M., Kurihara, C., Tada, Y., Tsou, T.L., van Rooijen, N., Lawton, M.T., Young, W.L., Liang, E.I., Nuki, Y., et al. (2011). Critical roles of macrophages in the formation of intracranial aneurysm. Stroke 42, 173-178.

147. Aoki, T., Kataoka, H., Morimoto, M., Nozaki, K., and Hashimoto, N. (2007). Macrophage-derived matrix metalloproteinase-2 and -9 promote the progression of cerebral aneurysms in rats. Stroke 38, 162-169.

148. Tronc, F., Mallat, Z., Lehoux, S., Wassef, M., Esposito, B., and Tedgui, A. (2000). Role of matrix metalloproteinases in blood flow-induced arterial enlargement: interaction with NO. Arterioscler Thromb Vasc Biol 20, E120-126.

149. Jin, D., Sheng, J., Yang, X., and Gao, B. (2007). Matrix metalloproteinases and tissue inhibitors of metalloproteinases expression in human cerebral ruptured and unruptured aneurysm. Surg Neurol 68 Suppl 2, S11-16; discussion S16.

150. Jayaraman, T., Paget, A., Shin, Y.S., Li, X., Mayer, J., Chaudhry, H., Niimi, Y., Silane, M., and Berenstein, A. (2008). TNF-alpha-mediated inflammation in cerebral aneurysms: a potential link to growth and rupture. Vasc Health Risk Manag 4, 805-817.

151. Kurki, M.I., Hakkinen, S.K., Frosen, J., Tulamo, R., von und zu Fraunberg, M., Wong, G., Tromp, G., Niemela, M., Hernesniemi, J., Jaaskelainen, J.E., et al. (2011). Upregulated signaling pathways in ruptured human saccular intracranial aneurysm wall: an emerging regulative role of Toll-like receptor signaling and nuclear factor-kappaB, hypoxia-inducible factor-1A, and ETS transcription factors. Neurosurgery 68, 1667-1675; discussion 1675-1666.

152. Loeys, B.L., Schwarze, U., Holm, T., Callewaert, B.L., Thomas, G.H., Pannu, H., De Backer, J.F., Oswald, G.L., Symoens, S., Manouvrier, S., et al. (2006). Aneurysm syndromes caused by mutations in the TGF-beta receptor. N Engl J Med 355, 788-798.

153. Sprague, A.H., and Khalil, R.A. (2009). Inflammatory cytokines in vascular dysfunction and vascular disease. Biochem Pharmacol 78, 539-552.

154. Frosen, J., Piippo, A., Paetau, A., Kangasniemi, M., Niemela, M., Hernesniemi, J., and Jaaskelainen, J. (2006). Growth factor receptor expression and remodeling of saccular cerebral artery aneurysm walls: implications for biological therapy preventing rupture. Neurosurgery 58, 534-541; discussion 534-541.

155. Ishibashi, R., Aoki, T., Nishimura, M., Hashimoto, N., and Miyamoto, S. (2010). Contribution of mast cells to cerebral aneurysm formation. Curr Neurovasc Res 7, 113-124.

156. Edlow, J.A. (2005). Diagnosis of subarachnoid hemorrhage. Neurocrit Care 2, 99-109.

157. Mason, A.M., Cawley, C.M., and Barrow, D.L. (2009). Surgical management of intracranial aneurysms in the endovascular era : review article. J Korean Neurosurg Soc 45, 133-142.

158. Molyneux, A., Kerr, R., Stratton, I., Sandercock, P., Clarke, M., Shrimpton, J., Holman, R., and International Subarachnoid Aneurysm Trial Collaborative, G. (2002). International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised trial. Lancet 360, 1267-1274.

159. Smith, T.R., Cote, D.J., Dasenbrock, H.H., Hamade, Y.J., Zammar, S.G., El Tecle, N.E., Batjer, H.H., and Bendok, B.R. (2015). Comparison of the Efficacy and Safety of Endovascular Coiling Versus Microsurgical Clipping for Unruptured Middle Cerebral Artery Aneurysms: A Systematic Review and Meta-Analysis. World Neurosurg 84, 942-953.

160. Ronkainen, A., Hernesniemi, J., Puranen, M., Niemitukia, L., Vanninen, R., Ryynanen, M., Kuivaniemi, H., and Tromp, G. (1997). Familial intracranial aneurysms. Lancet 349, 380-384.

161. Rinkel, G.J. (2005). Intracranial aneurysm screening: indications and advice for practice. Lancet Neurol 4, 122-128.

162. Wills, S., Ronkainen, A., van der Voet, M., Kuivaniemi, H., Helin, K., Leinonen, E., Frosen, J., Niemela, M., Jaaskelainen, J., Hernesniemi, J., et al. (2003). Familial intracranial aneurysms: an analysis of 346 multiplex Finnish families. Stroke 34, 1370-1374.

163. Nahed, B.V., Seker, A., Guclu, B., Ozturk, A.K., Finberg, K., Hawkins, A.A., DiLuna, M.L., State, M., Lifton, R.P., and Gunel, M. (2005). Mapping a Mendelian form of intracranial aneurysm to 1p34.3-p36.13. Am J Hum Genet 76, 172-179.

164. Farnham, J.M., Camp, N.J., Neuhausen, S.L., Tsuruda, J., Parker, D., MacDonald, J., and Cannon-Albright, L.A. (2004). Confirmation of chromosome 7q11 locus for predisposition to intracranial aneurysm. Hum Genet 114, 250-255.

165. Ozturk, A.K., Nahed, B.V., Bydon, M., Bilguvar, K., Goksu, E., Bademci, G., Guclu, B., Johnson, M.H., Amar, A., Lifton, R.P., et al. (2006). Molecular genetic analysis of two large kindreds with intracranial aneurysms demonstrates linkage to 11q24-25 and 14q23-31. Stroke 37, 1021-1027.

166. Ruigrok, Y.M., Wijmenga, C., Rinkel, G.J., van't Slot, R., Baas, F., Wolfs, M., Westerveld, A., and Roos, Y.B. (2008). Genomewide linkage in a large Dutch family with intracranial aneurysms: replication of 2 loci for intracranial aneurysms to chromosome 1p36.11-p36.13 and Xp22.2-p22.32. Stroke 39, 1096-1102.

167. Olson, J.M., Vongpunsawad, S., Kuivaniemi, H., Ronkainen, A., Hernesniemi, J., Ryynanen, M., Kim, L.L., and Tromp, G. (2002). Search for intracranial aneurysm susceptibility gene(s) using Finnish families. BMC Med Genet 3, 7.

168. Onda, H., Kasuya, H., Yoneyama, T., Takakura, K., Hori, T., Takeda, J., Nakajima, T., and Inoue, I. (2001). Genomewide-linkage and haplotype-association studies map intracranial aneurysm to chromosome 7q11. Am J Hum Genet 69, 804-819.

169. Kim, C.J., Park, S.S., Lee, H.S., Chung, H.J., Choi, W., Chung, J.H., Kim, J.M., and Hong, S.T. (2011). Identification of an autosomal dominant locus for intracranial aneurysm through a model-based family collection in a geographically limited area. J Hum Genet 56, 464-466.

170. Yamada, S., Utsunomiya, M., Inoue, K., Nozaki, K., Inoue, S., Takenaka, K., Hashimoto, N., and Koizumi, A. (2004). Genome-wide scan for Japanese familial intracranial aneurysms: linkage to several chromosomal regions. Circulation 110, 3727-3733.

171. Mineharu, Y., Inoue, K., Inoue, S., Yamada, S., Nozaki, K., Hashimoto, N., and Koizumi, A. (2007). Model-based linkage analyses confirm chromosome 19q13.3 as a susceptibility locus for intracranial aneurysm. Stroke 38, 1174-1178.

172. Verlaan, D.J., Dube, M.P., St-Onge, J., Noreau, A., Roussel, J., Satge, N., Wallace, M.C., and Rouleau, G.A. (2006). A new locus for autosomal dominant intracranial aneurysm, ANIB4, maps to chromosome 5p15.2-14.3. J Med Genet 43, e31.

173. Santiago-Sim, T., Depalma, S.R., Ju, K.L., McDonough, B., Seidman, C.E., Seidman, J.G., and Kim, D.H. (2009). Genomewide linkage in a large Caucasian family maps a new locus for intracranial aneurysms to chromosome 13q. Stroke 40, S57-60.

174. Foroud, T., Sauerbeck, L., Brown, R., Anderson, C., Woo, D., Kleindorfer, D., Flaherty, M.L., Deka, R., Hornung, R., Meissner, I., et al. (2008). Genome screen to detect linkage to intracranial aneurysm susceptibility genes: the Familial Intracranial Aneurysm (FIA) study. Stroke 39, 1434-1440.

175. Liu, J., Pho, R.W., Kour, A.K., Zhang, A.H., and Ong, B.K. (1997). Neurologic deficit and recovery in the donor limb following cross-C7 transfer in brachial-plexus injury. J Reconstr Microsurg 13, 237-242; discussion 242-233.

176. Huttunen, T., von und zu Fraunberg, M., Frosen, J., Lehecka, M., Tromp, G., Helin, K., Koivisto, T., Rinne, J., Ronkainen, A., Hernesniemi, J., et al. (2010). Saccular intracranial aneurysm disease: distribution of site, size, and age suggests different etiologies for aneurysm formation and rupture in 316 familial and 1454 sporadic eastern Finnish patients. Neurosurgery 66, 631-638; discussion 638.

177. Broderick, J.P., Brown, R.D., Jr., Sauerbeck, L., Hornung, R., Huston, J., 3rd, Woo, D., Anderson, C., Rouleau, G., Kleindorfer, D., Flaherty, M.L., et al. (2009). Greater rupture risk for familial as compared to sporadic unruptured intracranial aneurysms. Stroke 40, 1952-1957.

178. Akagawa, H., Tajima, A., Sakamoto, Y., Krischek, B., Yoneyama, T., Kasuya, H., Onda, H., Hori, T., Kubota, M., Machida, T., et al. (2006). A haplotype spanning two genes, ELN and LIMK1, decreases their transcripts and confers susceptibility to intracranial aneurysms. Hum Mol Genet 15, 1722-1734.

179. Alg, V.S., Sofat, R., Houlden, H., and Werring, D.J. (2013). Genetic risk factors for intracranial aneurysms: a meta-analysis in more than 116,000 individuals. Neurology 80, 2154-2165.

180. Krischek, B., Tajima, A., Akagawa, H., Narita, A., Ruigrok, Y., Rinkel, G., Wijmenga, C., Feigl, G.C., Kim, C.J., Hori, T., et al. (2010). Association of the Jun dimerization protein 2 gene with intracranial aneurysms in Japanese and Korean cohorts as compared to a Dutch cohort. Neuroscience 169, 339-343.

181. Weinsheimer, S., Goddard, K.A., Parrado, A.R., Lu, Q., Sinha, M., Lebedeva, E.R., Ronkainen, A., Niemela, M., Khusnutdinova, E.K., Khusainova, R.I., et al. (2007). Association of kallikrein gene polymorphisms with intracranial aneurysms. Stroke 38, 2670-2676.

182. Semmler, A., Linnebank, M., Krex, D., Gotz, A., Moskau, S., Ziegler, A., and Simon, M. (2008). Polymorphisms of homocysteine metabolism are associated with intracranial aneurysms. Cerebrovasc Dis 26, 425-429.

183. Low, S.K., Zembutsu, H., Takahashi, A., Kamatani, N., Cha, P.C., Hosono, N., Kubo, M., Matsuda, K., and Nakamura, Y. (2011). Impact of LIMK1, MMP2 and TNF-alpha variations for intracranial aneurysm in Japanese population. J Hum Genet 56, 211-216.

184. Chen, Z., Ma, J., Cen, Y., Liu, Y., and You, C. (2013). The angiotensin converting enzyme insertion/deletion polymorphism and intracranial aneurysm: a meta-analysis of case-control studies. Neurol India 61, 293-299.

185. Inoue, K., Mineharu, Y., Inoue, S., Yamada, S., Matsuda, F., Nozaki, K., Takenaka, K., Hashimoto, N., and Koizumi, A. (2006). Search on chromosome 17 centromere reveals TNFRSF13B as a susceptibility gene for intracranial aneurysm: a preliminary study. Circulation 113, 2002-2010.

186. Weinsheimer, S., Lenk, G.M., van der Voet, M., Land, S., Ronkainen, A., Alafuzoff, I., Kuivaniemi, H., and Tromp, G. (2007). Integration of expression profiles and genetic mapping data to identify candidate genes in intracranial aneurysm. Physiol Genomics 32, 45-57.

187. Krischek, B., Kasuya, H., Tajima, A., Akagawa, H., Sasaki, T., Yoneyama, T., Ujiie, H., Kubo, O., Bonin, M., Takakura, K., et al. (2008). Network-based gene expression analysis of intracranial aneurysm tissue reveals role of antigen presenting cells. Neuroscience 154, 1398-1407.

188. Shi, C., Awad, I.A., Jafari, N., Lin, S., Du, P., Hage, Z.A., Shenkar, R., Getch, C.C., Bredel, M., Batjer, H.H., et al. (2009). Genomics of human intracranial aneurysm wall. Stroke 40, 1252-1261.

189. Yu, L., Fan, J., Wang, S., Zhang, D., Wang, R., Zhao, Y., and Zhao, J. (2014). Gene expression profiles in intracranial aneurysms. Neurosci Bull 30, 99-106.

190. Pera, J., Korostynski, M., Krzyszkowski, T., Czopek, J., Slowik, A., Dziedzic, T., Piechota, M., Stachura, K., Moskala, M., Przewlocki, R., et al. (2010). Gene expression profiles in human ruptured and unruptured intracranial aneurysms: what is the role of inflammation? Stroke 41, 224-231.

191. Marchese, E., Vignati, A., Albanese, A., Nucci, C.G., Sabatino, G., Tirpakova, B., Lofrese, G., Zelano, G., and Maira, G. (2010). Comparative evaluation of genome-wide gene expression profiles in ruptured and unruptured human intracranial aneurysms. J Biol Regul Homeost Agents 24, 185-195.

192. Nakaoka, H., Tajima, A., Yoneyama, T., Hosomichi, K., Kasuya, H., Mizutani, T., and Inoue, I. (2014). Gene expression profiling reveals distinct molecular signatures associated with the rupture of intracranial aneurysm. Stroke 45, 2239-2245.

193. Chen, L., Wan, J.Q., Zhou, J.P., Fan, Y.L., and Jiang, J.Y. (2013). Gene expression analysis of ruptured and un-ruptured saccular intracranial aneurysm. Eur Rev Med Pharmacol Sci 17, 1374-1381.

194. Li, L., Yang, X., Jiang, F., Dusting, G.J., and Wu, Z. (2009). Transcriptome-wide characterization of gene expression associated with unruptured intracranial aneurysms. Eur Neurol 62, 330-337.

195. Sabatino, G., Rigante, L., Minella, D., Novelli, G., Della Pepa, G.M., Esposito, G., Albanese, A., Maira, G., and Marchese, E. (2013). Transcriptional profile characterization for the identification of peripheral blood biomarkers in patients with cerebral aneurysms. J Biol Regul Homeost Agents 27, 729-738.

196. Jin, H., Li, C., Ge, H., Jiang, Y., and Li, Y. (2013). Circulating microRNA: a novel potential biomarker for early diagnosis of intracranial aneurysm rupture a case control study. J Transl Med 11, 296.

197. Jiang, Y., Zhang, M., He, H., Chen, J., Zeng, H., Li, J., and Duan, R. (2013). MicroRNA/mRNA profiling and regulatory network of intracranial aneurysm. BMC Med Genomics 6, 36.

198. Liu, D., Han, L., Wu, X., Yang, X., Zhang, Q., and Jiang, F. (2014). Genome-wide microRNA changes in human intracranial aneurysms. BMC Neurol 14, 188.

199. Kleinloog, R., Verweij, B.H., van der Vlies, P., Deelen, P., Swertz, M.A., de Muynck, L., Van Damme, P., Giuliani, F., Regli, L., van der Zwan, A., et al. (2016). RNA Sequencing Analysis of Intracranial Aneurysm Walls Reveals Involvement of Lysosomes and Immunoglobulins in Rupture. Stroke 47, 1286-1293.

200. Roder, C., Kasuya, H., Harati, A., Tatagiba, M., Inoue, I., and Krischek, B. (2012). Meta-analysis of microarray gene expression studies on intracranial aneurysms. Neuroscience 201, 105-113.

201. Zheng, X., Xue, C., Luo, G., Hu, Y., Luo, W., and Sun, X. (2015). Identification of crucial genes in intracranial aneurysm based on weighted gene coexpression network analysis. Cancer Gene Ther 22, 238-245.

202. Fan, J., Sun, W., Lin, M., Yu, K., Wang, J., Duan, D., Zheng, B., Yang, Z., and Wang, Q. (2016). Genetic association study identifies a functional CNV in the WWOX gene contributes to the risk of intracranial aneurysms. Oncotarget.

203. Lee, S., Kim, I.K., Ahn, J.S., Woo, D.C., Kim, S.T., Song, S., Koh, G.Y., Kim, H.S., Jeon, B.H., and Kim, I. (2015). Deficiency of endothelium-specific transcription factor Sox17 induces intracranial aneurysm. Circulation 131, 995-1005.

204. Pera, J., Korostynski, M., Golda, S., Piechota, M., Dzbek, J., Krzyszkowski, T., Dziedzic, T., Moskala, M., Przewlocki, R., Szczudlik, A., et al. (2013). Gene expression profiling of

blood in ruptured intracranial aneurysms: in search of biomarkers. J Cereb Blood Flow Metab 33, 1025-1031.

205. Tang, Y., Xu, H., Du, X., Lit, L., Walker, W., Lu, A., Ran, R., Gregg, J.P., Reilly, M., Pancioli, A., et al. (2006). Gene expression in blood changes rapidly in neutrophils and monocytes after ischemic stroke in humans: a microarray study. J Cereb Blood Flow Metab 26, 1089-1102.

206. Weinsheimer, S.M., Xu, H., Achrol, A.S., Stamova, B., McCulloch, C.E., Pawlikowska, L., Tian, Y., Ko, N.U., Lawton, M.T., Steinberg, G.K., et al. (2011). Gene expression profiling of blood in brain arteriovenous malformation patients. Transl Stroke Res 2, 575-587.

207. Bilguvar, K., Yasuno, K., Niemela, M., Ruigrok, Y.M., von Und Zu Fraunberg, M., van Duijn, C.M., van den Berg, L.H., Mane, S., Mason, C.E., Choi, M., et al. (2008). Susceptibility loci for intracranial aneurysm in European and Japanese populations. Nat Genet 40, 1472-1477.

208. Akiyama, K., Narita, A., Nakaoka, H., Cui, T., Takahashi, T., Yasuno, K., Tajima, A., Krischek, B., Yamamoto, K., Kasuya, H., et al. (2010). Genome-wide association study to identify genetic variants present in Japanese patients harboring intracranial aneurysms. J Hum Genet 55, 656-661.

209. Pasmant, E., Sabbagh, A., Vidaud, M., and Bieche, I. (2011). ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. FASEB J 25, 444-448.

210. Yasuno, K., Bakircioglu, M., Low, S.K., Bilguvar, K., Gaal, E., Ruigrok, Y.M., Niemela, M., Hata, A., Bijlenga, P., Kasuya, H., et al. (2011). Common variant near the endothelin receptor type A (EDNRA) gene is associated with intracranial aneurysm risk. Proc Natl Acad Sci U S A 108, 19707-19712.

211. Low, S.K., Takahashi, A., Cha, P.C., Zembutsu, H., Kamatani, N., Kubo, M., and Nakamura, Y. (2012). Genome-wide association study for intracranial aneurysm in the Japanese population identifies three candidate susceptible loci and a functional genetic variant at EDNRA. Hum Mol Genet 21, 2102-2110.

212. Gaal, E.I., Salo, P., Kristiansson, K., Rehnstrom, K., Kettunen, J., Sarin, A.P., Niemela, M., Jula, A., Raitakari, O.T., Lehtimaki, T., et al. (2012). Intracranial aneurysm risk locus 5q23.2 is associated with elevated systolic blood pressure. PLoS Genet 8, e1002563.

213. van 't Hof, F.N., Kurki, M.I., Kleinloog, R., de Bakker, P.I., von und zu Fraunberg, M., Jaaskelainen, J.E., Gaal, E.I., Lehto, H., Kivisaari, R., Laakso, A., et al. (2014). Genetic risk load according to the site of intracranial aneurysms. Neurology 83, 34-39.

214. Foroud, T., Koller, D.L., Lai, D., Sauerbeck, L., Anderson, C., Ko, N., Deka, R., Mosley, T.H., Fornage, M., Woo, D., et al. (2012). Genome-wide association study of intracranial aneurysms confirms role of Anril and SOX17 in disease risk. Stroke 43, 2846-2852.

215. Foroud, T., Lai, D., Koller, D., Van't Hof, F., Kurki, M.I., Anderson, C.S., Brown, R.D., Jr., Connolly, E.S., Eriksson, J.G., Flaherty, M., et al. (2014). Genome-wide association study of intracranial aneurysm identifies a new association on chromosome 7. Stroke 45, 3194-3199.

216. Matarin, M., Brown, W.M., Scholz, S., Simon-Sanchez, J., Fung, H.C., Hernandez, D., Gibbs, J.R., De Vrieze, F.W., Crews, C., Britton, A., et al. (2007). A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. Lancet Neurol 6, 414-420.

217. Abrantes, P., Santos, M.M., Sousa, I., Xavier, J.M., Francisco, V., Krug, T., Sobral, J., Matos, M., Martins, M., Jacinto, A., et al. (2015). Genetic Variants Underlying Risk of Intracranial Aneurysms: Insights from a GWAS in Portugal. PLoS One 10, e0133422.

218. Yan, J., Hitomi, T., Takenaka, K., Kato, M., Kobayashi, H., Okuda, H., Harada, K.H., and Koizumi, A. (2015). Genetic study of intracranial aneurysms. Stroke 46, 620-626.

219. Farlow, J.L., Lin, H., Sauerbeck, L., Lai, D., Koller, D.L., Pugh, E., Hetrick, K., Ling, H., Kleinloog, R., van der Vlies, P., et al. (2015). Lessons learned from whole exome sequencing in multiplex families affected by a complex genetic disorder, intracranial aneurysm. PLoS One 10, e0121104.

220. Kenny, E.E., Pe'er, I., Karban, A., Ozelius, L., Mitchell, A.A., Ng, S.M., Erazo, M., Ostrer, H., Abraham, C., Abreu, M.T., et al. (2012). A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. PLoS Genet 8, e1002559.

221. Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., et al. (2010). The genome-wide structure of the Jewish people. Nature 466, 238-242.

222. Sidore, C., Busonero, F., Maschio, A., Porcu, E., Naitza, S., Zoledziewska, M., Mulas, A., Pistis, G., Steri, M., Danjou, F., et al. (2015). Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nat Genet 47, 1272-1281.

223. Sanna, S., Pitzalis, M., Zoledziewska, M., Zara, I., Sidore, C., Murru, R., Whalen, M.B., Busonero, F., Maschio, A., Costa, G., et al. (2010). Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. Nat Genet 42, 495-497.

224. Dhandapany, P.S., Razzaque, M.A., Muthusami, U., Kunnoth, S., Edwards, J.J., Mulero-Navarro, S., Riess, I., Pardo, S., Sheng, J., Rani, D.S., et al. (2014). RAF1 mutations in childhood-onset dilated cardiomyopathy. Nat Genet 46, 635-639.

225. Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. Nature 461, 489-494.

226. Gessner, B.D., Gillingham, M.B., Birch, S., Wood, T., and Koeller, D.M. (2010). Evidence for an association between infant mortality and a carnitine palmitoyltransferase 1A genetic variant. Pediatrics 126, 945-951.

227. Anand, S.S., Yusuf, S., Jacobs, R., Davis, A.D., Yi, Q., Gerstein, H., Montague, P.A., and Lonn, E. (2001). Risk factors, atherosclerosis, and cardiovascular disease among Aboriginal people in Canada: the Study of Health Assessment and Risk Evaluation in Aboriginal Peoples (SHARE-AP). Lancet 358, 1147-1153.

228. Noel, M., Dewailly, E., Chateau-Degat, M.L., Counil, E., Laouan-Sidi, E.A., and Lonn, E. (2012). Cardiovascular risk factors and subclinical atherosclerosis among Nunavik Inuit. Atherosclerosis 221, 558-564.

229. Counil, E., Julien, P., Lamarche, B., Chateau-Degat, M.L., Ferland, A., and Dewailly, E. (2009). Association between trans-fatty acids in erythrocytes and pro-atherogenic lipid profiles among Canadian Inuit of Nunavik: possible influences of sex and age. Br J Nutr 102, 766-776.

230. Arbour, L., Asuri, S., Whittome, B., Polanco, F., and Hegele, R.A. (2015). The Genetics of Cardiovascular Disease in Canadian and International Aboriginal Populations. Can J Cardiol 31, 1094-1115.

231. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19, 1655-1664.

232. Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., Reynolds, A., Ostrer, H., Mezey, J.G., and Bustamante, C.D. (2012). PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Hum Biol 84, 343-364.

233. Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford, T.W., Jr., Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature 506, 225-229.

234. Rasmussen, M., Sikora, M., Albrechtsen, A., Korneliussen, T.S., Moreno-Mayar, J.V., Poznik, G.D., Zollikofer, C.P., Ponce de Leon, M.S., Allentoft, M.E., Moltke, I., et al. (2015). The ancestry and affiliations of Kennewick Man. Nature 523, 455-458.

235. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8, e1002967.

236. Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329, 75-78.

237. Raghavan, M., Steinrucken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Avila-Arcos, M.C., Malaspinas, A.S., et al. (2015). POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science 349, aab3884.

238. Duforet-Frebourg, N., Luu, K., Laval, G., Bazin, E., and Blum, M.G. (2016). Detecting Genomic Signatures of Natural Selection with Principal Component Analysis: Application to the 1000 Genomes Data. Mol Biol Evol 33, 1082-1093.

239. Collins, H.B. (1956). Vanished Mystery Men of Hudson-Bay. Natl Geogr 110, 669-687.

240. G, M.-R. (1980). Qitdlarssuaq, l'histoire d'une migration polaire.(Montreal: Presses de l'Universite de Montreal).

241. Perestenko, P.V., Pooler, A.M., Noorbakhshnia, M., Gray, A., Bauccio, C., and Jeffrey McIlhinney, R.A. (2010). Copines-1, -2, -3, -6 and -7 show different calcium-dependent intracellular membrane translocation and targeting. FEBS J 277, 5174-5189.

242. Lee, J.Y., Kim, G., Park, S., Kang, S.M., Jang, Y., and Lee, S.H. (2015). Associations between Genetic Variants and Angiographic Characteristics in Patients with Coronary Artery Disease. J Atheroscler Thromb 22, 363-371.

243. Ley, K., Miller, Y.I., and Hedrick, C.C. (2011). Monocyte and macrophage dynamics during atherogenesis. Arterioscler Thromb Vasc Biol 31, 1506-1516.

244. Fan, X.J., Zhao, H.D., Yu, G., Zhong, X.L., Yao, H., and Yang, Q.D. (2015). Role of inflammatory responses in the pathogenesis of human cerebral aneurysm. Genet Mol Res 14, 9062-9070.

245. Fukami, K., Yamagishi, S., and Okuda, S. (2014). Role of AGEs-RAGE system in cardiovascular disease. Curr Pharm Des 20, 2395-2402.

246. Anbarasan, C., Bavanilatha, M., Latchumanadhas, K., and Ajit Mullasari, S. (2015). ICAM-1 molecular mechanism and genome wide SNP's association studies. Indian Heart J 67, 282-287.

247. Leung, S., Qureshi, S.A., Kerr, I.M., Darnell, J.E., Jr., and Stark, G.R. (1995). Role of STAT2 in the alpha interferon signaling pathway. Mol Cell Biol 15, 1312-1317.

248. Khorooshi, R., Babcock, A.A., and Owens, T. (2008). NF-kappaB-driven STAT2 and CCL2 expression in astrocytes in response to brain injury. J Immunol 181, 7284-7291.

249. Lagor, W.R., Fields, D.W., Bauer, R.C., Crawford, A., Abt, M.C., Artis, D., Wherry, E.J., and Rader, D.J. (2014). Genetic manipulation of the ApoF/Stat2 locus supports an important role for type I interferon signaling in atherosclerosis. Atherosclerosis 233, 234-241.

250. Szelag, M., Piaszyk-Borychowska, A., Plens-Galaska, M., Wesoly, J., and Bluyssen, H.A. (2016). Targeted inhibition of STATs and IRFs as a potential treatment strategy in cardiovascular disease. Oncotarget.

251. Alavi, A., Hood, J.D., Frausto, R., Stupack, D.G., and Cheresh, D.A. (2003). Role of Raf in vascular protection from distinct apoptotic stimuli. Science 301, 94-96.

252. Deng, Y., Larrivee, B., Zhuang, Z.W., Atri, D., Moraes, F., Prahst, C., Eichmann, A., and Simons, M. (2013). Endothelial RAF1/ERK activation regulates arterial morphogenesis. Blood 121, 3988-3996, S3981-3989.

253. Timofeeva, A.V., Goryunova, L.E., Khaspekov, G.L., Kovalevskii, D.A., Scamrov, A.V., Bulkina, O.S., Karpov, Y.A., Talitskii, K.A., Buza, V.V., Britareva, V.V., et al. (2006). Altered gene expression pattern in peripheral blood leukocytes from patients with arterial hypertension. Ann N Y Acad Sci 1091, 319-335.

254. Ayari, H., Legedz, L., Cerutti, C., Lantelme, P., Feugier, P., Gustin, M.P., Lohez, O., Nehme, A., Li, J.Y., Gharbi-Chihi, J., et al. (2014). Mutual amplification of corticosteroids and angiotensin systems in human vascular smooth muscle cells and carotid atheroma. J Mol Med (Berl) 92, 1201-1208.

255. Ebbesson, S.O., Voruganti, V.S., Higgins, P.B., Fabsitz, R.R., Ebbesson, L.O., Laston, S., Harris, W.S., Kennish, J., Umans, B.D., Wang, H., et al. (2015). Fatty acids linked to cardiovascular mortality are associated with risk factors. Int J Circumpolar Health 74, 28055.

256. Synowiec, E., Wojcik, K.A., Wojcik, R., Wisniewski, K., Fila, M., Tokarz, P., Bienkowski, M., Jaskolski, D., and Blasiak, J. (2016). Expression and variability of lipid metabolism genes in intracranial aneurysm. Cell Mol Biol (Noisy-le-grand) 62, 73-82.

257. Bigner, D.D., McLendon, R.E., and Bruner, J.M. (2006). Russell and Rubinstein's pathology of tumors of the nervous system.(London

New York, NY: Hodder Arnold ;

Distributed in the United States of America by Oxford University Press).

258. McDonald, D.A., Shi, C., Shenkar, R., Gallione, C.J., Akers, A.L., Li, S., De Castro, N., Berg, M.J., Corcoran, D.L., Awad, I.A., et al. (2014). Lesions from patients with sporadic cerebral cavernous malformations harbor somatic mutations in the CCM genes: evidence for a common biochemical pathway for CCM pathogenesis. Hum Mol Genet 23, 4357-4370.

259. Crose, L.E., Hilder, T.L., Sciaky, N., and Johnson, G.L. (2009). Cerebral cavernous malformation 2 protein promotes smad ubiquitin regulatory factor 1-mediated RhoA degradation in endothelial cells. J Biol Chem 284, 13301-13305.

260. Zhou, Z., Tang, A.T., Wong, W.Y., Bamezai, S., Goddard, L.M., Shenkar, R., Zhou, S., Yang, J., Wright, A.C., Foley, M., et al. (2016). Cerebral cavernous malformations arise from endothelial gain of MEKK3-KLF2/4 signalling. Nature 532, 122-126.

261. Cullere, X., Plovie, E., Bennett, P.M., MacRae, C.A., and Mayadas, T.N. (2015). The cerebral cavernous malformation proteins CCM2L and CCM2 prevent the activation of the MAP kinase MEKK3. Proc Natl Acad Sci U S A 112, 14284-14289.

262. Fisher, O.S., Deng, H., Liu, D., Zhang, Y., Wei, R., Deng, Y., Zhang, F., Louvi, A., Turk, B.E., Boggon, T.J., et al. (2015). Structure and vascular function of MEKK3-cerebral cavernous malformations 2 complex. Nat Commun 6, 7937.

263. Kleaveland, B., Zheng, X., Liu, J.J., Blum, Y., Tung, J.J., Zou, Z., Sweeney, S.M., Chen, M., Guo, L., Lu, M.M., et al. (2009). Regulation of cardiovascular development and integrity by the heart of glass-cerebral cavernous malformation protein pathway. Nat Med 15, 169-176.

264. Macek Jilkova, Z., Lisowska, J., Manet, S., Verdier, C., Deplano, V., Geindreau, C., Faurobert, E., Albiges-Rizo, C., and Duperray, A. (2014). CCM proteins control endothelial beta1 integrin dependent response to shear stress. Biol Open 3, 1228-1235.

265. Hogan, B.M., Bussmann, J., Wolburg, H., and Schulte-Merker, S. (2008). ccm1 cell autonomously regulates endothelial cellular morphogenesis and vascular tubulogenesis in zebrafish. Hum Mol Genet 17, 2424-2432.

266. Can, A., and Du, R. (2016). Association of Hemodynamic Factors With Intracranial Aneurysm Formation and Rupture: Systematic Review and Meta-analysis. Neurosurgery 78, 510-520.

267. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303.

268. Fedorova, S.A., Reidla, M., Metspalu, E., Metspalu, M., Rootsi, S., Tambets, K., Trofimova, N., Zhadanov, S.I., Hooshiar Kashani, B., Olivieri, A., et al. (2013). Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. BMC Evol Biol 13, 127.

269. Yunusbayev, B., Metspalu, M., Jarve, M., Kutuev, I., Rootsi, S., Metspalu, E., Behar, D.M., Varendi, K., Sahakyan, H., Khusainova, R., et al. (2012). The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. Mol Biol Evol 29, 359-365.

270. Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., Akhmetova, V., Balanovska, E., Balanovsky, O., Turdikulova, S., et al. (2015). The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. PLoS Genet 11, e1005068.

271. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature 526, 68-74.

272. Allentoft, M.E., Sikora, M., Sjogren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlstrom, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature 522, 167-172.

273. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7.

274. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet 2, e190.

275. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 81, 1084-1097.

276. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402-408.

277. Schievink, W.I. (1997). Genetics of intracranial aneurysms. Neurosurgery 40, 651-662; discussion 662-653.

278. Colonna, V., Pagani, L., Xue, Y., and Tyler-Smith, C. (2011). A world in a grain of sand: human history from genetic data. Genome biology 12, 234.

279. Patricia Gadsby, L.S. (2004). The Inuit Paradox In DISCOVER. (

280. Jogl, G., Hsiao, Y.S., and Tong, L. (2004). Structure and function of carnitine acyltransferases. Ann N Y Acad Sci 1033, 17-29.

281. Murthy, M.S., and Pande, S.V. (1987). Malonyl-CoA binding site and the overt carnitine palmitoyltransferase activity reside on the opposite sides of the outer mitochondrial membrane. Proc Natl Acad Sci U S A 84, 378-382.

282. van der Leij, F.R., Huijkman, N.C., Boomsma, C., Kuipers, J.R., and Bartelds, B. (2000). Genomics of the human carnitine acyltransferase genes. Mol Genet Metab 71, 139-153.

283. Bennett, M.J., and Santani, A.B. (1993). Carnitine Palmitoyltransferase 1A Deficiency. In GeneReviews, R.A. Pagon, M.P. Adam, T.D. Bird, C.R. Dolan, C.T. Fong, andK. Stephens, eds. (Seattle (WA).

284. Wieser, T. (1993). Carnitine Palmitoyltransferase II Deficiency. In GeneReviews, R.A. Pagon, M.P. Adam, T.D. Bird, C.R. Dolan, C.T. Fong, andK. Stephens, eds. (Seattle (WA).

285. Orngreen, M.C., Duno, M., Ejstrup, R., Christensen, E., Schwartz, M., Sacchetti, M., and Vissing, J. (2005). Fuel utilization in subjects with carnitine palmitoyltransferase 2 gene mutations. Annals of neurology 57, 60-66.

286. Proust, F., Lucas, M., and Dewailly, E. (2014). Fatty acid profiles among the Inuit of Nunavik: current status and temporal change. Prostaglandins, leukotrienes, and essential fatty acids 90, 159-167.

287. Labonte, M.E., Dewailly, E., Lucas, M., Chateau-Degat, M.L., Couture, P., and Lamarche, B. (2014). Traditional dietary pattern is associated with elevated cholesterol among the Inuit of Nunavik. Journal of the Academy of Nutrition and Dietetics 114, 1208-1215 e1203.

288. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

289. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research 38, e164.

290. Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56-65.

291. (2014). Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP). In. (Seattle, WA

292. Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics 14, 178-192.

293. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81, 559-575.

294. Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263-265.

295. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-Statistics for the Analysis of Population-Structure. Evolution 38, 1358-1370.

296. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nature methods 7, 248-249.

297. Schwarz, J.M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nature methods 7, 575-576.

298. Mak, C.M., Lam, C.W., Fong, N.C., Siu, W.K., Lee, H.C., Siu, T.S., Lai, C.K., Law, C.Y., Tong, S.F., Poon, W.T., et al. (2011). Fatal viral infection-associated encephalopathy in

two Chinese boys: a genetically determined risk factor of thermolabile carnitine palmitoyltransferase II variants. J Hum Genet 56, 617-621.

299. Muoio, D.M., Noland, R.C., Kovalik, J.P., Seiler, S.E., Davies, M.N., DeBalsi, K.L., Ilkayeva, O.R., Stevens, R.D., Kheterpal, I., Zhang, J., et al. (2012). Muscle-specific deletion of carnitine acetyltransferase compromises glucose tolerance and metabolic flexibility. Cell metabolism 15, 764-777.

300. Taggart, R.T., Smail, D., Apolito, C., and Vladutiu, G.D. (1999). Novel mutations associated with carnitine palmitoyltransferase II deficiency. Human mutation 13, 210-220.

301. Illsinger, S., Lucke, T., Peter, M., Ruiter, J.P., Wanders, R.J., Deschauer, M., Handig, I., Wuyts, W., and Das, A.M. (2008). Carnitine-palmitoyltransferase 2 deficiency: novel mutations and relevance of newborn screening. Am J Med Genet A 146A, 2925-2928.

302. Gessner, B.D., Gillingham, M.B., Johnson, M.A., Richards, C.S., Lambert, W.E., Sesser, D., Rien, L.C., Hermerath, C.A., Skeels, M.R., Birch, S., et al. (2011). Prevalence and distribution of the c.1436C-->T sequence variant of carnitine palmitoyltransferase 1A among Alaska Native infants. The Journal of pediatrics 158, 124-129.

303. Gobin, S., Bonnefont, J.P., Prip-Buus, C., Mugnier, C., Ferrec, M., Demaugre, F., Saudubray, J.M., Rostane, H., Djouadi, F., Wilcox, W., et al. (2002). Organization of the human liver carnitine palmitoyltransferase 1 gene (CPT1A) and identification of novel mutations in hypoketotic hypoglycaemia. Hum Genet 111, 179-189.

304. Lemas, D.J., Wiener, H.W., O'Brien, D.M., Hopkins, S., Stanhope, K.L., Havel, P.J., Allison, D.B., Fernandez, J.R., Tiwari, H.K., and Boyer, B.B. (2012). Genetic polymorphisms in carnitine palmitoyltransferase 1A gene are associated with variation in body composition and fasting lipid traits in Yup'ik Eskimos. J Lipid Res 53, 175-184.

305. Rajakumar, C., Ban, M.R., Cao, H., Young, T.K., Bjerregaard, P., and Hegele, R.A. (2009). Carnitine palmitoyltransferase IA polymorphism P479L is common in Greenland Inuit and is associated with elevated plasma apolipoprotein A-I. J Lipid Res 50, 1223-1228.

306. Macaulay, A., Orr, P., Macdonald, S., Elliott, L., Brown, R., Durcan, A., and Martin, B. (2004). Mortality in the Kivalliq Region of Nunavut, 1987-1996. Int J Circumpolar Health 63 Suppl 2, 80-85.

307. Miyagawa, T., Kawashima, M., Nishida, N., Ohashi, J., Kimura, R., Fujimoto, A., Shimada, M., Morishita, S., Shigeta, T., Lin, L., et al. (2008). Variant between CPT1B and CHKB associated with susceptibility to narcolepsy. Nat Genet 40, 1324-1328.

308. Miyagawa, T., Miyadera, H., Tanaka, S., Kawashima, M., Shimada, M., Honda, Y., Tokunaga, K., and Honda, M. (2011). Abnormally low serum acylcarnitine levels in narcolepsy patients. Sleep 34, 349-353A.

309. Chen, Y., Mizuguchi, H., Yao, D., Ide, M., Kuroda, Y., Shigematsu, Y., Yamaguchi, S., Yamaguchi, M., Kinoshita, M., and Kido, H. (2005). Thermolabile phenotype of carnitine palmitoyltransferase II variations as a predisposing factor for influenza-associated encephalopathy. FEBS letters 579, 2040-2044.

310. Yao, D., Mizuguchi, H., Yamaguchi, M., Yamada, H., Chida, J., Shikata, K., and Kido, H. (2008). Thermal instability of compound variants of carnitine palmitoyltransferase II and impaired mitochondrial fuel utilization in influenza-associated encephalopathy. Human mutation 29, 718-727.

311. Stehbens, W.E. (1983). The Pathology of Intracranial Arterial Aneurysms and Their Complications. In Intracranial Aneurysms, J.L. Fox, ed. (Springer-Verlag New York), pp 272-357.

312. Rinkel, G.J. (2008). Natural history, epidemiology and screening of unruptured intracranial aneurysms. J Neuroradiol 35, 99-103.

313. Qureshi, A.I., Suri, M.F., Nasar, A., Kirmani, J.F., Divani, A.A., He, W., and Hopkins, L.N. (2005). Trends in hospitalization and mortality for subarachnoid hemorrhage and unruptured aneurysms in the United States. Neurosurgery 57, 1-8; discussion 1-8.

314. Feigin, V.L., Lawes, C.M., Bennett, D.A., Barker-Collo, S.L., and Parag, V. (2009). Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. Lancet Neurol 8, 355-369.

315. Teunissen, L.L., Rinkel, G.J., Algra, A., and van Gijn, J. (1996). Risk factors for subarachnoid hemorrhage: a systematic review. Stroke 27, 544-549.

316. Brown, R.D., Jr., Wiebers, D.O., and Forbes, G.S. (1990). Unruptured intracranial aneurysms and arteriovenous malformations: frequency of intracranial hemorrhage and relationship of lesions. J Neurosurg 73, 859-863.

317. Clare, C.E., and Barrow, D.L. (1992). Infectious intracranial aneurysms. Neurosurg Clin N Am 3, 551-566.

318. Benoit, B.G., and Wortzman, G. (1973). Traumatic cerebral aneurysms. Clinical features and natural history. J Neurol Neurosurg Psychiatry 36, 127-138.

319. Krischek, B., and Inoue, I. (2006). The genetics of intracranial aneurysms. J Hum Genet 51, 587-594.

320. Yasuno, K., Bilguvar, K., Bijlenga, P., Low, S.K., Krischek, B., Auburger, G., Simon, M., Krex, D., Arlier, Z., Nayak, N., et al. (2010). Genome-wide association study of intracranial aneurysm identifies three new risk loci. Nat Genet 42, 420-425.

321. Halal, F., Mohr, G., Toussi, T., and Napoleon Martinez, S. (1983). Intracranial aneurysms: a report of a large pedigree. Am J Med Genet 15, 89-95.

322. Mathieu, J., Perusse, L., Allard, P., Prevost, C., Cantin, L., Bouchard, J.M., and DeBraekeleer, M. (1996). Epidemiological study of reptured intracranial aneurysms in the Saguenay-Lac-Saint-Jean region (Quebec, Canada). Can J Neurol Sci 23, 184-188.

323. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

324. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol 6, e1001025.

325. Price, A.L., Kryukov, G.V., de Bakker, P.I., Purcell, S.M., Staples, J., Wei, L.J., and Sunyaev, S.R. (2010). Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 86, 832-838.

326. Wang, G.T., Peng, B., and Leal, S.M. (2014). Variant association tools for quality control and analysis of large-scale sequence and genotyping array data. Am J Hum Genet 94, 770-783.

327. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589-595.

328. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol 29, 24-26.

329. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904-909.

330. Eichstaedt, C.A., Antao, T., Pagani, L., Cardona, A., Kivisild, T., and Mormina, M. (2014). The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. PLoS One 9, e93314.

331. Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.X., Antao, T., Capredon, M., Sambo, C., Radimilahy, C., Rakotoarisoa, J.A., Blench, R.M., et al. (2014). Genome-wide evidence of Austronesian-Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proc Natl Acad Sci U S A 111, 936-941.

332. Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J.D., and Lin, X. (2013). Sequence kernel association tests for the combined effect of rare and common variants. Am J Hum Genet 92, 841-853.

333. Excoffier, L., and Lischer, H.E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10, 564-567.

334. Liu, W., Morito, D., Takashima, S., Mineharu, Y., Kobayashi, H., Hitomi, T., Hashikata, H., Matsuura, N., Yamazaki, S., Toyoda, A., et al. (2011). Identification of RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development. PLoS One 6, e22542.

335. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature methods 6, 343-345.

336. Morito, D., Nishikawa, K., Hoseki, J., Kitamura, A., Kotani, Y., Kiso, K., Kinjo, M., Fujiyoshi, Y., and Nagata, K. (2014). Moyamoya disease-associated protein mysterin/RNF213 is a novel AAA+ ATPase, which dynamically changes its oligomeric state. Sci Rep 4, 4442.

337. Lindner, C., Sigruner, A., Walther, F., Bogdahn, U., Couraud, P.O., Schmitz, G., and Schlachetzki, F. (2012). ATP-binding cassette transporters in immortalised human brain microvascular endothelial cells in normal and hypoxic conditions. Exp Transl Stroke Med 4, 9.

338. Kamada, F., Aoki, Y., Narisawa, A., Abe, Y., Komatsuzaki, S., Kikuchi, A., Kanno, J., Niihori, T., Ono, M., Ishii, N., et al. (2011). A genome-wide association study identifies RNF213 as the first Moyamoya disease gene. J Hum Genet 56, 34-40.

339. Herve, D., Philippi, A., Belbouab, R., Zerah, M., Chabrier, S., Collardeau-Frachon, S., Bergametti, F., Essongue, A., Berrou, E., Krivosic, V., et al. (2014). Loss of alpha1beta1 soluble guanylate cyclase, the major nitric oxide receptor, leads to moyamoya and achalasia. Am J Hum Genet 94, 385-394.

340. Shoemaker, L.D., Clark, M.J., Patwardhan, A., Chandratillake, G., Garcia, S., Chen, R., Morgan, A.A., Leng, N., Kirk, S., Chen, R., et al. (2016). Disease Variant Landscape of a Large Multiethnic Population of Moyamoya Patients by Exome Sequencing. G3 (Bethesda) 6, 41-49.

341. Hitomi, T., Habu, T., Kobayashi, H., Okuda, H., Harada, K.H., Osafune, K., Taura, D., Sone, M., Asaka, I., Ameku, T., et al. (2013). Downregulation of Securin by the variant RNF213 R4810K (rs112735431, G>A) reduces angiogenic activity of induced pluripotent stem cell-derived vascular endothelial cells from moyamoya patients. Biochem Biophys Res Commun 438, 13-19.

342. Kobayashi, H., Matsuda, Y., Hitomi, T., Okuda, H., Shioi, H., Matsuda, T., Imai, H., Sone, M., Taura, D., Harada, K.H., et al. (2015). Biochemical and Functional Characterization of RNF213 (Mysterin) R4810K, a Susceptibility Mutation of Moyamoya Disease, in Angiogenesis In Vitro and In Vivo. J Am Heart Assoc 4.

343. Hitomi, T., Habu, T., Kobayashi, H., Okuda, H., Harada, K.H., Osafune, K., Taura, D., Sone, M., Asaka, I., Ameku, T., et al. (2013). The moyamoya disease susceptibility variant RNF213 R4810K (rs112735431) induces genomic instability by mitotic abnormality. Biochem Biophys Res Commun 439, 419-426.

344. Sonobe, S., Fujimura, M., Niizuma, K., Nishijima, Y., Ito, A., Shimizu, H., Kikuchi, A., Arai-Ichinoi, N., Kure, S., and Tominaga, T. (2014). Temporal profile of the vascular anatomy evaluated by 9.4-T magnetic resonance angiography and histopathological analysis in mice lacking RNF213: a susceptibility gene for moyamoya disease. Brain Res 1552, 64-71.

345. Sonobe, S., Fujimura, M., Niizuma, K., Fujimura, T., Furudate, S., Nishijima, Y., Kure, S., and Tominaga, T. (2014). Increased vascular MMP-9 in mice lacking RNF213: moyamoya disease susceptibility gene. Neuroreport 25, 1442-1446.

346. Ito, A., Fujimura, M., Niizuma, K., Kanoke, A., Sakata, H., Morita-Fujimura, Y., Kikuchi, A., Kure, S., and Tominaga, T. (2015). Enhanced post-ischemic angiogenesis in mice lacking RNF213; a susceptibility gene for moyamoya disease. Brain Res 1594, 310-320.

347. Ohkubo, K., Sakai, Y., Inoue, H., Akamine, S., Ishizaki, Y., Matsushita, Y., Sanefuji, M., Torisu, H., Ihara, K., Sardiello, M., et al. (2015). Moyamoya disease susceptibility gene RNF213 links inflammatory and angiogenic signals in endothelial cells. Sci Rep 5, 13191.

348. Fukushima, Y., Miyawaki, S., Inoue, T., Shimizu, S., Yoshikawa, G., Imai, H., Saito, N., and Tsutsumi, K. (2015). Repeated de novo aneurysm formation after anastomotic surgery: Potential risk of genetic variant RNF213 c.14576G>A. Surg Neurol Int 6, 41.

349. Cecchi, A.C., Guo, D., Ren, Z., Flynn, K., Santos-Cortez, R.L., Leal, S.M., Wang, G.T., Regalado, E.S., Steinberg, G.K., Shendure, J., et al. (2014). RNF213 rare variants in an ethnically diverse population with Moyamoya disease. Stroke 45, 3200-3207.

350. Kiando, S.R., Barlassina, C., Cusi, D., Galan, P., Lathrop, M., Plouin, P.F., Jeunemaitre, X., and Bouatia-Naji, N. (2015). Exome sequencing in seven families and gene-based association studies indicate genetic heterogeneity and suggest possible candidates for fibromuscular dysplasia. J Hypertens 33, 1802-1810.

351. Koizumi, A., Kobayashi, H., Liu, W., Fujii, Y., Senevirathna, S.T., Nanayakkara, S., Okuda, H., Hitomi, T., Harada, K.H., Takenaka, K., et al. (2013). P.R4810K, a polymorphism of RNF213, the susceptibility gene for moyamoya disease, is associated with blood pressure. Environ Health Prev Med 18, 121-129.

352. Miyawaki, S., Imai, H., Shimizu, M., Yagi, S., Ono, H., Mukasa, A., Nakatomi, H., Shimizu, T., and Saito, N. (2013). Genetic variant RNF213 c.14576G>A in various phenotypes of intracranial major artery stenosis/occlusion. Stroke 44, 2894-2897.

353. Miyawaki, S., Imai, H., Takayanagi, S., Mukasa, A., Nakatomi, H., and Saito, N. (2012). Identification of a genetic variant common to moyamoya disease and intracranial major artery stenosis/occlusion. Stroke 43, 3371-3374.

354. Smith, K.R., Leventer, R.J., Mackay, M.T., Pope, K., Gillies, G., Delatycki, M.B., Amor, D.J., Bahlo, M., and Lockhart, P.J. (2014). Identification of a novel RNF213 variant in a family with heterogeneous intracerebral vasculopathy. Int J Stroke 9, E26-27.

355. Bang, O.Y., Ryoo, S., Kim, S.J., Yoon, C.H., Cha, J., Yeon, J.Y., Kim, K.H., Kim, G.M., Chung, C.S., Lee, K.H., et al. (2015). Adult Moyamoya Disease: A Burden of Intracranial Stenosis in East Asians? PLoS One 10, e0130663.

356. Wu, Z., Jiang, H., Zhang, L., Xu, X., Zhang, X., Kang, Z., Song, D., Zhang, J., Guan, M., and Gu, Y. (2012). Molecular analysis of RNF213 gene for moyamoya disease in the Chinese Han population. PLoS One 7, e48179.

357. Moteki, Y., Onda, H., Kasuya, H., Yoneyama, T., Okada, Y., Hirota, K., Mukawa, M., Nariai, T., Mitani, S., and Akagawa, H. (2015). Systematic Validation of RNF213 Coding Variants in Japanese Patients With Moyamoya Disease. J Am Heart Assoc 4.

358. Lee, M.J., Chen, Y.F., Fan, P.C., Wang, K.C., Wang, K., Wang, J., and Kuo, M.F. (2015). Mutation genotypes of RNF213 gene from moyamoya patients in Taiwan. J Neurol Sci 353, 161-165.

359. Koizumi, A., Kobayashi, H., Hitomi, T., Harada, K.H., Habu, T., and Youssefian, S. (2016). A new horizon of moyamoya disease and associated health risks explored through RNF213. Environ Health Prev Med 21, 55-70.

360. Scott, R.M., and Smith, E.R. (2009). Moyamoya disease and moyamoya syndrome. N Engl J Med 360, 1226-1237.

361. Kessler, K., Borges, L.F., Ho-Tin-Noe, B., Jondeau, G., Michel, J.B., and Vranckx, R. (2014). Angiogenesis and remodelling in human thoracic aortic aneurysms. Cardiovasc Res 104, 147-159.

362. Hoh, B.L., Hosaka, K., Downes, D.P., Nowicki, K.W., Wilmer, E.N., Velat, G.J., and Scott, E.W. (2014). Stromal cell-derived factor-1 promoted angiogenesis and inflammatory cell infiltration in aneurysm walls. J Neurosurg 120, 73-86.

363. Li, B., Li, F., Chi, L., Zhang, L., and Zhu, S. (2013). The expression of SPARC in human intracranial aneurysms and its relationship with MMP-2/-9. PLoS One 8, e58490.

364. Imai, Y., Sato, K., Ishikawa, T., Comerford, A., David, T., and Yamaguchi, T. (2010). ATP transport in saccular cerebral aneurysms at arterial bends. Ann Biomed Eng 38, 927-934.

365. Power, A.S., Pham, T., Loiselle, D.S., Crossman, D.H., Ward, M.L., and Hickey, A.J. (2016). Impaired ADP channeling to mitochondria and elevated reactive oxygen species in hypertensive hearts. Am J Physiol Heart Circ Physiol 310, H1649-1657.

366. Vanhoutte, P.M., Zhao, Y., Xu, A., and Leung, S.W. (2016). Thirty Years of Saying NO: Sources, Fate, Actions, and Misfortunes of the Endothelium-Derived Vasodilator Mediator. Circ Res 119, 375-396.

367. Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39, 906-913.

368. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42, 565-569.

369. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88, 76-82.

370. Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A., Lee, S.H., Robinson, M.R., Perry, J.R., Nolte, I.M., van Vliet-Ostaptchouk, J.V., et al. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet 47, 1114-1120.

371. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89, 82-93.

372. Laird, N.M., Horvath, S., and Xu, X. (2000). Implementing a unified approach to family-based tests of association. Genet Epidemiol 19 Suppl 1, S36-42.

373. Barnes, L.D., Garrison, P.N., Siprashvili, Z., Guranowski, A., Robinson, A.K., Ingram, S.W., Croce, C.M., Ohta, M., and Huebner, K. (1996). Fhit, a putative tumor suppressor in humans, is a dinucleoside 5',5'''-P1,P3-triphosphate hydrolase. Biochemistry 35, 11529-11535.

374. Karras, J.R., Paisie, C.A., and Huebner, K. (2014). Replicative Stress and the FHIT Gene: Roles in Tumor Suppression, Genome Stability and Prevention of Carcinogenesis. Cancers (Basel) 6, 1208-1219.

375. Starke, R.M., Chalouhi, N., Ali, M.S., Jabbour, P.M., Tjoumakaris, S.I., Gonzalez, L.F., Rosenwasser, R.H., Koch, W.J., and Dumont, A.S. (2013). The role of oxidative stress in cerebral aneurysm formation and rupture. Curr Neurovasc Res 10, 247-255.

376. Nikpay, M., Seda, O., Tremblay, J., Petrovich, M., Gaudet, D., Kotchen, T.A., Cowley, A.W., Jr., and Hamet, P. (2012). Genetic mapping of habitual substance use, obesity-related traits, responses to mental and physical stress, and heart rate and blood pressure measurements reveals shared genes that are overrepresented in the neural synapse. Hypertens Res 35, 585-591.

377. Chemnitz, J.M., Lanfranco, A.R., Braunstein, I., and Riley, J.L. (2006). B and T lymphocyte attenuator-mediated signal transduction provides a potent inhibitory signal to primary human CD4 T cells that can be initiated by multiple phosphotyrosine motifs. J Immunol 176, 6603-6614.

378. Steinberg, M.W., Huang, Y., Wang-Zhu, Y., Ware, C.F., Cheroutre, H., and Kronenberg, M. (2013). BTLA interaction with HVEM expressed on CD8(+) T cells promotes survival and memory generation in response to a bacterial infection. PLoS One 8, e77992.

379. Cheng, T., Bai, J., Chung, C.S., Chen, Y., Biron, B.M., and Ayala, A. (2016). Enhanced Innate Inflammation Induced by Anti-BTLA Antibody in Dual Insult Model of Hemorrhagic Shock/Sepsis. Shock 45, 40-49.

380. Yoo, B.H., Zagryazhskaya, A., Li, Y., Koomson, A., Khan, I.A., Sasazuki, T., Shirasawa, S., and Rosen, K.V. (2015). Upregulation of ATG3 contributes to autophagy induced by the detachment of intestinal epithelial cells from the extracellular matrix, but promotes autophagy-independent apoptosis of the attached cells. Autophagy 11, 1230-1246.

381. Bharath, L.P., Mueller, R., Li, Y., Ruan, T., Kunz, D., Goodrich, R., Mills, T., Deeter, L., Sargsyan, A., Anandh Babu, P.V., et al. (2014). Impairment of autophagy in endothelial cells prevents shear-stress-induced increases in nitric oxide bioavailability. Can J Physiol Pharmacol 92, 605-612.

382. Ding, L., Abebe, T., Beyene, J., Wilke, R.A., Goldberg, A., Woo, J.G., Martin, L.J., Rothenberg, M.E., Rao, M., Hershey, G.K., et al. (2013). Rank-based genome-wide analysis reveals the association of ryanodine receptor-2 gene variants with childhood asthma among human populations. Hum Genomics 7, 16.

383. Wang, G.R., Surks, H.K., Tang, K.M., Zhu, Y., Mendelsohn, M.E., and Blanton, R.M. (2013). Steroid-sensitive gene 1 is a novel cyclic GMP-dependent protein kinase I substrate in vascular smooth muscle cells. J Biol Chem 288, 24972-24983.

384. Tremblay, F., Huard, C., Dow, J., Gareski, T., Will, S., Richard, A.M., Syed, J., Bailey, S., Brenneman, K.A., Martinez, R.V., et al. (2012). Loss of coiled-coil domain containing 80 negatively modulates glucose homeostasis in diet-induced obese mice. Endocrinology 153, 4290-4303.

385. Jarrin, M., Pandit, T., and Gunhaga, L. (2012). A balance of FGF and BMP signals regulates cell cycle exit and Equarin expression in lens cells. Mol Biol Cell 23, 3266-3274.

386. Song, X., Sato, Y., Sekiguchi, K., Tanaka, H., and Ohta, K. (2013). Equarin is involved in cell adhesion by means of heparan sulfate proteoglycan during lens development. Dev Dyn 242, 23-29.

387. Ferraro, A., Schepis, F., Leone, V., Federico, A., Borbone, E., Pallante, P., Berlingieri, M.T., Chiappetta, G., Monaco, M., Palmieri, D., et al. (2013). Tumor suppressor role of the CL2/DRO1/CCDC80 gene in thyroid carcinogenesis. J Clin Endocrinol Metab 98, 2834-2843.

388. Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS Genet 6, e1001154.

389. Karavitakis, E., Kitsiou-Tzeli, S., Xaidara, A., Kosma, K., Makrythanasis, P., Apazidou, E., Kanavakis, E., and Tzetis, M. (2014). Microduplication 3q13.2q13.31 identified in a male with dysmorphic features and multiple congenital anomalies. Am J Med Genet A 164A, 666-670.

390. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48, 1279-1283.

391. Delaneau, O., Marchini, J., and Zagury, J.F. (2012). A linear complexity phasing method for thousands of genomes. Nature methods 9, 179-181.

392. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336-2337.

393. San Lucas, F.A., Wang, G., Scheet, P., and Peng, B. (2012). Integrated annotation and analysis of genetic variants from next-generation sequencing studies with variant tools. Bioinformatics 28, 421-422.

394. Henn, B.M., Botigue, L.R., Bustamante, C.D., Clark, A.G., and Gravel, S. (2015). Estimating the mutation load in human genomes. Nat Rev Genet 16, 333-343.

395. Oh, H.J., Choung, H.W., Lee, H.K., Park, S.J., Lee, J.H., Lee, D.S., Seo, B.M., and Park, J.C. (2015). CPNE7, a preameloblast-derived factor, regulates odontoblastic differentiation of mesenchymal stem cells. Biomaterials 37, 208-217.

396. Atsumi, T., Singh, R., Sabharwal, L., Bando, H., Meng, J., Arima, Y., Yamada, M., Harada, M., Jiang, J.J., Kamimura, D., et al. (2014). Inflammation amplifier, a new paradigm in cancer biology. Cancer Res 74, 8-14.

397. Ruiz-Larranaga, O., Migliorini, P., Uribarri, M., Czirjak, L., Alcaro, M.C., Del Amo, J., Iriondo, M., Manzano, C., Escorza-Trevino, S., and Estonba, A. (2016). Genetic association study of systemic lupus erythematosus and disease subphenotypes in European populations. Clin Rheumatol 35, 1161-1168.

398. Tsutsumi, S., Ogino, I., Miyajima, M., Ikeda, T., Shindo, N., Yasumoto, Y., Ito, M., and Arai, H. (2013). Genomic causes of multiple cerebral cavernous malformations in a Japanese population. J Clin Neurosci 20, 667-669.

399. Fujimura, M., Sonobe, S., Nishijima, Y., Niizuma, K., Sakata, H., Kure, S., and Tominaga, T. (2014). Genetics and Biomarkers of Moyamoya Disease: Significance of RNF213 as a Susceptibility Gene. J Stroke 16, 65-72.

400. Manousaki, D., Kent, J.W., Jr., Haack, K., Zhou, S., Xie, P., Greenwood, C.M., Brassard, P., Newman, D.E., Cole, S., Umans, J.G., et al. (2016). Toward Precision Medicine: TBC1D4

Disruption Is Common Among the Inuit and Leads to Underdiagnosis of Type 2 Diabetes. Diabetes Care.

401. Williamson, S.H., Hubisz, M.J., Clark, A.G., Payseur, B.A., Bustamante, C.D., and Nielsen, R. (2007). Localizing recent adaptive evolution in the human genome. PLoS Genet 3, e90.

402. Dudley, J.T., Kim, Y., Liu, L., Markov, G.J., Gerold, K., Chen, R., Butte, A.J., and Kumar, S. (2012). Human genomic disease variants: a neutral evolutionary explanation. Genome Res 22, 1383-1394.

403. Sudmant, P.H., Mallick, S., Nelson, B.J., Hormozdiari, F., Krumm, N., Huddleston, J., Coe, B.P., Baker, C., Nordenfelt, S., Bamshad, M., et al. (2015). Global diversity, population stratification, and selection of human copy-number variation. Science 349, aab3761.

404. Lopopolo, M., Borsting, C., Pereira, V., and Morling, N. (2016). A study of the peopling of Greenland using next generation sequencing of complete mitochondrial genomes. Am J Phys Anthropol.

405. Sasagawa, S., Nishimura, Y., Sawada, H., Zhang, E., Okabe, S., Murakami, S., Ashikawa, Y., Yuge, M., Kawaguchi, K., Kawase, R., et al. (2016). Comparative Transcriptome Analysis Identifies CCDC80 as a Novel Gene Associated with Pulmonary Arterial Hypertension. Front Pharmacol 7, 142.

406. Sazzini, M., Gnecchi Ruscone, G.A., Giuliani, C., Sarno, S., Quagliariello, A., De Fanti, S., Boattini, A., Gentilini, D., Fiorito, G., Catanoso, M., et al. (2016). Complex interplay between neutral and adaptive evolution shaped differential genomic background and disease susceptibility along the Italian peninsula. Sci Rep 6, 32513.

407. Simonti, C.N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D.S., Chisholm, R.L., Crosslin, D.R., Hebbring, S.J., Jarvik, G.P., Kullo, I.J., et al. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. Science 351, 737-741.

408. International Consortium for Blood Pressure Genome-Wide Association, S., Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., et al. (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478, 103-109.

409. Pagenstecher, A., Stahl, S., Sure, U., and Felbor, U. (2009). A two-hit mechanism causes cerebral cavernous malformations: complete inactivation of CCM1, CCM2 or CCM3 in affected endothelial cells. Hum Mol Genet 18, 911-918.

410. Ayturk, U.M., Couto, J.A., Hann, S., Mulliken, J.B., Williams, K.L., Huang, A.Y., Fishman, S.J., Boyd, T.K., Kozakewich, H.P., Bischoff, J., et al. (2016). Somatic Activating Mutations in GNAQ and GNA11 Are Associated with Congenital Hemangioma. Am J Hum Genet 98, 1271.

411. Castillo, S.D., Tzouanacou, E., Zaw-Thin, M., Berenjeno, I.M., Parker, V.E., Chivite, I., Mila-Guasch, M., Pearce, W., Solomon, I., Angulo-Urarte, A., et al. (2016). Somatic activating mutations in Pik3ca cause sporadic venous malformations in mice and humans. Sci Transl Med 8, 332ra343.

412. Macmurdo, C.F., Wooderchak-Donahue, W., Bayrak-Toydemir, P., Le, J., Wallenstein, M.B., Milla, C., Teng, J.M., Bernstein, J.A., and Stevenson, D.A. (2016). RASA1 somatic mutation and variable expressivity in capillary malformation/arteriovenous malformation (CM/AVM) syndrome. Am J Med Genet A 170, 1450-1454.

413. Kato, N., Loh, M., Takeuchi, F., Verweij, N., Wang, X., Zhang, W., Kelly, T.N., Saleheen, D., Lehne, B., Mateo Leach, I., et al. (2015). Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. Nat Genet 47, 1282-1293.

414. Fagny, M., Patin, E., MacIsaac, J.L., Rotival, M., Flutre, T., Jones, M.J., Siddle, K.J., Quach, H., Harmant, C., McEwen, L.M., et al. (2015). The epigenomic landscape of African rainforest hunter-gatherers and farmers. Nat Commun 6, 10047.

415. Nuttle, X., Giannuzzi, G., Duyzend, M.H., Schraiber, J.G., Narvaiza, I., Sudmant, P.H., Penn, O., Chiatante, G., Malig, M., Huddleston, J., et al. (2016). Emergence of a Homo sapiens-specific gene family and chromosome 16p11.2 CNV susceptibility. Nature 536, 205-209.

416. (2015). Canadian Climate Normals or Averages 1981-2010. In. (Environment Canada, Environment Canada.

417. Bang, H.O., Dyerberg, J., and Hjoorne, N. (1976). The composition of food consumed by Greenland Eskimos. Acta medica Scandinavica 200, 69-73.

418. Middaugh, J.P. (1990). Cardiovascular deaths among Alaskan Natives, 1980-86. American journal of public health 80, 282-285.

419. Young, T.K., Moffatt, M.E., and O'Neil, J.D. (1993). Cardiovascular diseases in a Canadian Arctic population. American journal of public health 83, 881-887.

420. Fodor, J.G., Helis, E., Yazdekhasti, N., and Vohnout, B. (2014). "Fishing" for the origins of the "Eskimos and heart disease" story: facts or wishful thinking? Can J Cardiol 30, 864-868.

421. Bjerregaard, P., Young, T.K., and Hegele, R.A. (2003). Low incidence of cardiovascular disease among the Inuit--what is the evidence? Atherosclerosis 166, 351-357.

422. Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W., et al. (2009). Signals of recent positive selection in a worldwide sample of human populations. Genome Res 19, 826-837.

423. Mainland, J.D., Keller, A., Li, Y.R., Zhou, T., Trimmer, C., Snyder, L.L., Moberly, A.H., Adipietro, K.A., Liu, W.L., Zhuang, H., et al. (2014). The missense of smell: functional variability in the human odorant receptor repertoire. Nature neuroscience 17, 114-120.

424. Gilad, Y., and Lancet, D. (2003). Population differences in the human functional olfactory repertoire. Mol Biol Evol 20, 307-314.

425. Powell, L.M., Wallis, S.C., Pease, R.J., Edwards, Y.H., Knott, T.J., and Scott, J. (1987). A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell 50, 831-840.

426. Kendrick, J.S., Chan, L., and Higgins, J.A. (2001). Superior role of apolipoprotein B48 over apolipoprotein B100 in chylomicron assembly and fat absorption: an investigation of apobec-1 knock-out and wild-type mice. Biochem J 356, 821-827.

427. Takenaka, Y., Shimano, T., Mori, T., Hou, I.C., Ohinata, K., and Yoshikawa, M. (2008). Enterostatin reduces serum cholesterol levels by way of a CCK(1) receptor-dependent mechanism. Peptides 29, 2175-2178.

428. van Tilbeurgh, H., Bezzine, S., Cambillau, C., Verger, R., and Carriere, F. (1999). Colipase: structure and interaction with pancreatic lipase. Biochim Biophys Acta 1441, 173-184.

429. Davis, A.E., 3rd, Whitehead, A.S., Harrison, R.A., Dauphinais, A., Bruns, G.A., Cicardi, M., and Rosen, F.S. (1986). Human inhibitor of the first component of complement, C1: characterization of cDNA clones and localization of the gene to chromosome 11. Proc Natl Acad Sci U S A 83, 3161-3165.

430. Aulak, K.S., Davis, A.E., 3rd, Donaldson, V.H., and Harrison, R.A. (1993). Chymotrypsin inhibitory activity of normal C1-inhibitor and a P1 Arg to His mutant: evidence for the presence of overlapping reactive centers. Protein science : a publication of the Protein Society 2, 727-732.

431. Madden, C.J., Tupone, D., Cano, G., and Morrison, S.F. (2013). alpha2 Adrenergic receptor-mediated inhibition of thermogenesis. J Neurosci 33, 2017-2028.

432. Livingston, A., Low, J., and Morris, B. (1984). Effects of clonidine and xylazine on body temperature in the rat. Br J Pharmacol 81, 189-193.

433. Berlan, M., and Lafontan, M. (1985). Evidence that epinephrine acts preferentially as an antilipolytic agent in abdominal human subcutaneous fat cells: assessment by analysis of beta and alpha 2 adrenoceptor properties. European journal of clinical investigation 15, 341-348.

434. Lafontan, M., and Berlan, M. (1995). Fat cell alpha 2-adrenoceptors: the regulation of fat cell function and lipolysis. Endocrine reviews 16, 716-738.

435. Carpene, C., Berlan, M., and Lafontan, M. (1983). Influence of development and reduction of fat stores on the antilipolytic alpha 2-adrenoceptor in hamster adipocytes: comparison with adenosine and beta-adrenergic lipolytic responses. J Lipid Res 24, 766-774.

436. Dixen, K., Basse, A.L., Murholm, M., Isidor, M.S., Hansen, L.H., Petersen, M.C., Madsen, L., Petrovic, N., Nedergaard, J., Quistorff, B., et al. (2013). ERRgamma enhances UCP1 expression and fatty acid oxidation in brown adipocytes. Obesity 21, 516-524.

437. Alaynick, W.A., Kondo, R.P., Xie, W., He, W., Dufour, C.R., Downes, M., Jonker, J.W., Giles, W., Naviaux, R.K., Giguere, V., et al. (2007). ERRgamma directs and maintains the transition to oxidative metabolism in the postnatal heart. Cell metabolism 6, 13-24.

438. Consortium, G.T. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648-660.

439. Marchitti, S.A., Brocker, C., Orlicky, D.J., and Vasiliou, V. (2010). Molecular characterization, expression analysis, and role of ALDH3B1 in the cellular protection against oxidative stress. Free Radic Biol Med 49, 1432-1443.

440. Lo, C.M., Nordskog, B.K., Nauli, A.M., Zheng, S., Vonlehmden, S.B., Yang, Q., Lee, D., Swift, L.L., Davidson, N.O., and Tso, P. (2008). Why does the gut choose apolipoprotein B48 but not B100 for chylomicron formation? Am J Physiol Gastrointest Liver Physiol 294, G344-352.

441. Erlanson-Albertsson, C., and York, D. (1997). Enterostatin--a peptide regulating fat intake. Obes Res 5, 360-372.

442. D'Agostino, D., Cordle, R.A., Kullman, J., Erlanson-Albertsson, C., Muglia, L.J., and Lowe, M.E. (2002). Decreased postnatal survival and altered body weight regulation in procolipase-deficient mice. J Biol Chem 277, 7170-7177.

443. Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS biology 4, e72.

444. Gautier, M., and Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28, 1176-1177.

445. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913-918.